Hodge numbers of semistable representations

Derbyshire, Sam Luc

Awarding institution:
King's College London

Hodge numbers of semistable representations

Sam Luc Derbyshire

A thesis presented for the degree of Doctor of Philosophy,
under the supervision of Professor Fred Diamond.

King's College London

Faculty of Natural and Mathematical Sciences
Department of Mathematics

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text. The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

Abstract

Part I of this thesis concerns the relation, in p-adic Hodge theory, between the monodromy and the Hodge numbers of a filtered (φ, N)-module D. Studying the interaction of the Hodge and Newton polygons with N, we deduce that a monodromy operator of large rank forces the Hodge numbers of D to be large: this is the content of Theorem I.14.4.

This result can then be applied to various Galois representations. For instance, starting with a Hilbert modular form f over a totally real field F, it is known that we can attach to it a global p-adic Galois representation $\rho_f\colon \mathrm{G}_F \to \mathrm{GL}_2(E)$, for $E/\mathbb{Q}_p$ some finite extension. Choosing a prime $\mathfrak{p}$ of F above p, we can then study the local p-adic Galois representation $\rho_{f,\mathfrak{p}}\colon \mathrm{G}_{F_\mathfrak{p}} \to \mathrm{GL}_2(E)$. Assuming that $\rho_{f,\mathfrak{p}}$ is semistable with matching Hodge–Tate weights, we can then use Fontaine–Dieudonné theory to obtain a filtered (φ, N)-module $\mathrm{D}_{\mathrm{st}}(\rho_{f,\mathfrak{p}})$. Applying Theorem I.14.4, we deduce that if the weights of f are too small, then $\rho_{f,\mathfrak{p}}$ is in fact crystalline.

We also present in section I.16.1 an example of a non-split semistable non-crystalline extension of crystalline characters which is not "trivial by cyclotomic", even up to twists.

In part II, we explore parallel results on the automorphic side of the Langlands correspondence. Concentrating on the case of Hilbert modular forms, the approach is to study the p-adic integrality properties of Hecke operators. This naturally leads to the study of integral models of Hilbert modular varieties and of their associated automorphic vector bundles. A careful study of these leads to the introduction of certain renormalisation factors for the action of Hecke operators (Proposition II.5.2). We then prove the integrality of these renormalised Hecke operators using the q-expansion principle (Proposition II.5.4).

Finally, we use these integrality properties to arrive at conditions, dependent on the weight of a Hilbert modular form f, that guarantee that local components of f cannot be special; is the content of Theorem II.6.2. For instance, in the case that there is a unique prime $\mathfrak{p}$ above p in F, corresponding to the local component $\pi_{f,\mathfrak{p}}$ is a unique filtered $(\varphi, N, \mathrm{G}_{F_\mathfrak{p}})$-module $D_{f,\mathfrak{p}}$, and the results of Theorem I.14.4 and Theorem II.6.2 exactly match: if the weights ν_τ of f do not average at least 2, the former theorem shows that $D_{f,\mathfrak{p}}$ is potentially crystalline, while the latter shows that $\pi_{f,\mathfrak{p}}$ cannot be special. Other interesting behaviour can occur depending on the splitting behaviour of p in F, such as conditions on all pairs of weights of f as in the final example of section 6.2.1.

Acknowledgements

Mathematics in some sense has a common language: a language of symbols, technical definitions, computations, and logic. This language efficiently conveys some, but not all, modes of mathematical thinking. Mathematicians learn to translate certain things almost unconsciously from one mental mode to the other, so that some statements quickly become clear.
When the idea is clear, the formal setup is usually unnecessary and redundant. It's like a new toaster that comes with a 16-page manual. If you already understand toasters and if the toaster looks like previous toasters you've encountered, you might just plug it in and see if it works, rather than first reading all the details in the manual.
People familiar with ways of doing things in a subfield recognize various patterns of statements or formulas as idioms or circumlocution for certain concepts or mental images. But to people not already familiar with what's going on the same patterns are not very illuminating; they are often even misleading. ***The language is not alive except to those who use it.***

– William Thurston, *On Proof And Progress In Mathematics*

From the outside, mathematics can easily seem to be a technical, austere endeavour with little room for humanity. This largely results from the culture of mathematical communication, where an intuitive geometric idea can be obfuscated in writing to the length of several pages, rendering it impersonal. To compensate for this, one needs a steady source of intuitions and examples in order to reinvigorate the mathematics.
For this reason I have taken to heart the task of exposition, in particular regarding the geometric underpinnings of p-adic Hodge theory and the theory of automorphic forms. I must therefore thank my supervisor, Fred Diamond, for always generously sharing his intuitions and insights, helping me navigate ideas which might otherwise have seemed all too abstruse. Another inspiration has been Kevin Buzzard, whose talks always portrayed mathematics at their most earnest and genuine. I have also greatly benefited from the profuse advice and explanations I have received during my time at King's College London, in particular from Wansu Kim, James Newton and Shu Sasaki, who were all three gracious with their time and encouragement. I am also grateful to my examiners, Frazer Jarvis and Sarah Zerbes, for their helpful comments, suggestions and criticisms.
On a more personal note, I wish to thank my officemates, especially Tristan (for his levity, as well as his expertise concerning Nakayama's lemma) and Lambros (for his bon vivant and positive outlook), as well as my friends Callan, Cosmin, Maxime and Umberto for their unending support. Finally, I would be remiss not to express my gratitude to my parents for always looking out for me.

Contents

Introduction

The cohomological school of algebraic geometry, spearheaded by Grothendieck, proposes to study algebraic varieties through their cohomology. In roughly progressively increasing order of complexity as one goes down the diagram are the situations over the following fields:

$$
\begin{tikzcd}
 & & \mathbb{C} \arrow[dd] \\
\mathbb{F}_q \arrow[urr, dotted, no head] \arrow[dd] \arrow[dddr, dashed] & & \\
 & & \mathbb{C}((t)) \\
\mathbb{F}_q((t)) & & \\
 & \mathbb{Q}_p &
\end{tikzcd}
$$

These are all local situations, with the fields symbolically also representing their finite extensions (and, in the case of $\mathbb{C}$, also $\mathbb{R}$). There are also corresponding global situations: finite extensions of $\mathbb{R}(t)$, $\mathbb{F}_q(t)$ and $\mathbb{Q}$. In typical adélic fashion, these can be studied one prime at a time in order to land up in the simpler local case, with the understanding that there will be further global restrictions coming from reciprocity (the compatibility in behaviour between different places).
Here are some indications about the structure of the cohomology of algebraic varieties over these various fields:

- $\underline{\mathbb{C}}$ If the variety is smooth and projective, it can be viewed as a compact Kähler manifold, and Hodge theory furnishes its cohomology (say, singular cohomology with complex coefficients) with a Hodge structure. This is briefly summarised in section I.1. In more general circumstances, Deligne showed that one obtains instead a mixed Hodge structure [Deligne. Hodge III, Théorème 8.1.15, Proposition 8.2.2]. Given a $\mathbb{Q}$-vector space V, a mixed Hodge structure of weight i on V is the data of a Hodge filtration $F^\bullet$ on $V_\mathbb{C}$, and a weight filtration $W_\bullet$ on V such that the induced Hodge structures on the graded pieces $\mathrm{gr}_j^W V$ are *pure* Hodge structures of weight $i+j$, so that they look like the degree $i+j$ cohomology of a smooth proper complex algebraic variety. From the point of view of Hodge theory, then, the cohomology of arbitrary varieties looks like a successive extension of pure "smooth and proper" pieces, exhibited by $W_\bullet$. The construction of these mixed Hodge structures is briefly reviewed in section I.9.5.2.

- $\underline{\mathbb{F}_q}$ This is the situation that lead Weil to put forward the general study of cohomology theories of algebraic varieties. The idea was that, with a suitable cohomology theory for varieties over finite fields in hand, one could use a Lefschetz fixed point formula to count the number of points of such varieties. Such adapted cohomology theories were thus called Weil cohomology theories, or in the general case of not necessarily smooth, not necessarily proper varieties, Bloch–Ogus cohomology theories. The axioms for a Bloch–Ogus cohomology theory are similar to the Eilenberg–Steenrod axioms (including the dimension axiom; these are *ordinary* cohomology theories), yet account for the extra structure on the cohomology of algebraic varieties. Some examples are: singular cohomology of (the space of complex points of) complex varieties, ℓ-adic étale cohomology of varieties over fields of characteristic different from p, crystalline cohomology of varieties over fields of characteristic p.
  The cohomology of varieties over $\mathbb{F}_q$ can also be furnished with weights, mirroring the situation over $\mathbb{C}$. On top of this, the cohomology groups in question, by transport of structure, are also equipped with an action

of Gal $\left(\overline{\mathbb{F}_q}/\mathbb{F}_q\right)$, i.e. they have a Frobenius automorphism. For smooth proper varieties, the Weil conjectures (and in particular the Riemann hypothesis[1]) together with Poincaré duality lead us to expect the Frobenius endomorphism to take a restricted form: in degree i, the eigenvalues of the Frobenius endomorphism are *Weil numbers of weight* i. These are algebraic numbers α such that $|\iota(\alpha)| = q^{i/2}$ for any complex embedding $\iota\colon \overline{\mathbb{Q}} \to \mathbb{C}$. When all such eigenvalues of Frobenius are Weil numbers of weight i, we say the cohomology is *pure* of weight i. In a more general, mixed, situation, say for the ℓ-adic étale cohomology $\mathrm{H}^i_{\text{ét}}\left(X_{\overline{\mathbb{F}_q}}, \mathbb{Q}_\ell\right)$ of an arbitrary variety over $\mathbb{F}_q$, we expect instead to have a weight filtration $W_\bullet$ just as in the complex case, such that the j-th graded piece with respect to $W_\bullet$ is pure of weight $i + j$.

- $\underline{\mathbb{C}(\!(t)\!)}$ Here we can consider that we are dealing with the local behaviour around $t = 0$ of a family of complex varieties parametrised by t. The interesting behaviour occurs when we have a family of smooth proper varieties with exceptional singular special fibre (i.e. we have a smooth proper variety over $\mathbb{C}(\!(t)\!)$ that does not extend to a smooth model over $\mathbb{C}[\![t]\!]$). In this case the cohomology of the generic fibre acquires an action by a monodromy operator T, corresponding to the fundamental group of Spec $(\mathbb{C}(\!(t)\!))$. Using the theory of nearby cycles (section I.9.4), a limit mixed Hodge structure at $t = 0$ can be defined, and we can think of its weight filtration as being the monodromy filtration of the logarithm $N = \log(T)$ of the monodromy automorphism.
  This endomorphism N is a morphism of mixed Hodge structures of Hodge type $(-1, -1)$, which implies a bound on its index of nilpotence: if m is the longest length of a chain of non-zero Hodge numbers $\mathrm{h}^p, \mathrm{h}^{p-1}, \ldots, \mathrm{h}^{p-m+1}$, then $N^{m+1} = 0$. This is all part of the complex monodromy theorem, in this version from [Schmid. Variation, Theorem 6.1]. See section I.9.5.2.

- $\underline{L/\mathbb{F}_q(\!(t)\!)}$ In this case, for $\mathbb{Q}_\ell$-coefficients (with $\ell \neq p$), Deligne's relative version of the Riemann hypothesis [Deligne. Weil II, Théorème 3.3.1] applies. Starting with a preliminary result [Deligne. Weil II, Théorème 1.8.4], Ito proves [Ito. W-M in char. p, Theorem 1.1, Proposition 7.1] that the ℓ-adic étale cohomology $V = \mathrm{H}^i_{\text{ét}}(X_{\overline{L}}, \mathbb{Q}_\ell)$ of an arbitrary variety over L is equipped with a logarithm of monodromy endomorphism N whose monodromy filtration makes V mixed of weight i, i.e. $\mathrm{gr}^W_j V$ is pure of weight $i + j$ for every $j \in \mathbb{Z}$. This is the weight-monodromy theorem for varieties over a field of characteristic p.

- $\underline{K/\mathbb{Q}_p}$ Here again there are two different possible situations depending on the coefficients.
  - $\underline{\ell \neq p}$ If we consider an ℓ-adic coefficient field $E/\mathbb{Q}_\ell$ with $\ell \neq p$, we can use much of the same theory as the case for varieties over $\mathbb{F}_q$, with the analogy in mind $\mathbb{Q}_p \approx \mathbb{F}_p(\!(t)\!)$. In particular, we have Grothendieck's ℓ-adic monodromy theorem (Theorem I.9.1), which gives us a logarithm of monodromy N on the ℓ-adic étale cohomology $\mathrm{H}^i_{\text{ét}}(X_{\overline{K}}, \mathbb{Q}_\ell)$ of an arbitrary variety X over K. The monodromy–weight conjecture is then the statement that this is an ℓ-adic version of a mixed Hodge structure: the j-th graded piece with respect to the monodromy filtration of N should be pure of weight $i + j$.
  - $\underline{\ell = p}$ In this situation, the representation theory is much more complex. Essentially, Grothendieck's ℓ-adic monodromy theorem is a topological result that stems from the incompatibility of the respective topologies on G_K and E for $\ell \neq p$, restricting the possible continuous homomorphisms $\rho\colon \mathrm{G}_K \to \mathrm{GL}_n(E)$. However, in the situation of $E/\mathbb{Q}_p$, there are many more Galois representations, and it is no longer at all the case that every continuous representation $\rho\colon \mathrm{G}_K \to \mathrm{GL}_n(E)$ is described up to finite error by a nilpotent endomorphism like in the situation of unequal residue characteristics. Hodge theory has to be adapted to this situation: this is Fontaine's p-adic Hodge theory, which begins with the introduction of a field of p-adic periods $\mathbf{B}_{\mathrm{dR}}$ (section I.3) that plays the rôle in the p-adic world of $\mathbb{C}$ as a universal coefficient field. Replacing Grothendieck's ℓ-adic monodromy theorem in this context is the p-adic monodromy theorem of Berger, André, Kedlaya and Mebkhout (Theorem I.13.1). Furthermore, Fontaine proved that one can describe the relevant p-adic Galois representations fully

[1] Proved by Deligne for ℓ-adic étale cohomology [Deligne. Weil I, Théorème 1.6] [Deligne. Weil II, Théorème 3.3.1], and Kedlaya for crystalline cohomology [Kedlaya. Weil II, Theorem 6.6.2]. These two cases cover the behaviour for $\ell \neq p$ (respectively, $\ell = p$).

through judicious use of the usual suspects: a Hodge filtration $F^\bullet$, a Frobenius endomorphism φ, and a logarithm of monodromy N. This is the content of the main theorem of p-adic Hodge theory (Theorem I.12.1).

It is of crucial importance to investigate the interrelation of $F^\bullet$, φ, N and G_K in p-adic Hodge theory. The Galois action commutes with everything else; there are thus three relations left to consider.

- φ and N The relationship is given by the simple formula $N\varphi = p\varphi N$. The extra factor of p on the right hand side indicates that N should be considered as a morphism of the form $N\colon V \to V(-1)$, where (-1) denotes a Tate twist. It can be considered to be a manifestation of Griffiths transversality. This relation also pops up in the $\ell \neq p$ case, and is analogous to the factor of $2\pi\mathrm{i}$ appearing in Hodge theory over $\mathbb{C}$. In the context of p-adic Hodge theory, the formula follows from the structure of Fontaine's period ring $\mathbf{B}_{\mathrm{st}}$ (see section I.9.10, and definition I.11.1), and the expression of formula (I.9.9.1) describing the action of N on the cohomology of the appropriate geometric objects.

- φ and $F^\bullet$ This relation is best stated in terms of Newton polygons. In that language, the interaction is described by relative positions of Newton polygons: the Newton polygon of φ lies on or above that of $F^\bullet$. This is the subject of section I.8B.1, and underlies Fontaine's notion of weak admissibility (Definition I.8B.11) which is essential for the main theorem of p-adic Hodge theory (Theorem I.12.1).

- $F^\bullet$ and N Here we expect a bound on the nilpotence of N in terms of $F^\bullet$, mirroring the bound we noted above for varieties over $\mathbb{C}((t))$. In the p-adic situation, we prove an analogous statement as Theorem I.14.4, which imposes restrictions on N in terms of the Hodge filtration in a filtered (φ, N)-module.

A similar story can also be told in the automorphic world. According to Langlands' philosophy, objects of geometric origin (such as Galois representations afforded by the ℓ-adic étale cohomology of elliptic curves over $\mathbb{Q}$) should relate to objects of automorphic origin (such as modular forms for congruence subgroups of $\mathrm{SL}_2(\mathbb{Z})$). As an instance of Langlands' principle of functoriality, automorphic representations for various reductive algebraic groups over $\mathbb{Q}$ are brought into play.
Beyond modular forms (where one uses the reductive group $\mathrm{GL}_{2,\mathbb{Q}}$), one particular case of this conjectural correspondence stands out as being particularly amenable to study: the case of Hilbert modular forms over a totally real field F. The study of such objects brings about the reductive group $\boldsymbol{G} := \mathrm{Res}^F_{\mathbb{Q}}\mathrm{GL}_{2,F}$, as the analogues of modular curves in this context are Shimura varieties for $\boldsymbol{G}$.[2] These Hilbert modular varieties can be interpreted as moduli spaces of (polarisable) abelian varieties of dimension $d = [F : \mathbb{Q}]$, which are equipped with an action of $\mathcal{O}_F$. The presence of extra endomorphisms brings down the dimension of many related objects from d to 1 (or from $2d$ to 2), which strengthens the analogy with the case of modular forms (compared to, say, Siegel modular varieties, which correspond to the reductive group $\mathrm{GSp}_{2g,\mathbb{Q}}$).
One important subtlety arises, however. The modular curves $\mathrm{X}_0(N)$ have a dual purpose, as outlined in section II.1: they provide a geometric habitat for modular forms (which are seen as sections of certain modular vector bundles $\omega(k)$ and $\omega_0(k)$ over them), but also provide a way to realise Galois representations attached to modular forms [Deligne. $\rho_{f,\ell}$, (3.10)]. Indeed, Hodge theoretic principles yield the Shimura isomorphism [Deligne. $\rho_{f,\ell}$, Théorème 2.10]

$$\mathrm{H}^0\left(\mathrm{X}_0(N), \omega_0(k)\right) \oplus \overline{\mathrm{H}^0\left(\mathrm{X}_0(N), \omega_0(k)\right)} \xrightarrow{\sim} \mathrm{H}^1_{\mathrm{dR},!}\left(\mathrm{Y}_0(N)/\mathbb{C}, \mathcal{L}(k)\right),$$

for certain local systems $\mathcal{L}(k)$. This allows us to pick out, by using Hecke operators, a 2-dimensional ℓ-adic Galois representation of $\mathrm{G}_{\mathbb{Q}}$ associated with a modular form f of weight $k \geqslant 2$ within the parabolic ℓ-adic étale cohomology groups, endowed with $\mathrm{G}_{\mathbb{Q}}$-action,

$$\mathrm{H}^1_{\text{ét},!}\left(\mathrm{Y}_0(N)_{\bar{\mathbb{Q}}}, \mathcal{L}_\ell(k)\right),$$

[2] Or for the related group $\boldsymbol{G}^*$, as defined in section II.4.2.

for $\mathcal{L}_\ell(k)$ a certain lisse ℓ-adic sheaf of $\mathbb{Q}_\ell$-vector spaces related to $\mathcal{L}(k)$. The case of $k = 1$ can then be tackled by methods of congruences [Deligne–Serre, Théorème 4.1].

In trying to mimic this approach with Hilbert modular varieties such as $\mathrm{Y}_0(\mathfrak{n})$ (with $\mathfrak{n} \subseteq \mathcal{O}_F$ an ideal of $\mathcal{O}_F$), the same Hodge theoretic principles lead us to the study of

$$\mathrm{IH}^d_{\text{ét}}\left(\overline{\mathrm{Y}}_0(\mathfrak{n})_{\overline{\mathbb{Q}}}, \mathcal{L}_\ell\right).$$

This is the middle dimensional ℓ-adic étale intersection cohomology of the minimal compactification of $\mathrm{Y}_0(\mathfrak{n})$, with coefficients in certain lisse ℓ-adic sheaves $\mathcal{L}_\ell$ of $\mathbb{Q}_\ell$-vector spaces. These groups are a natural counterpart, in the setting of Hilbert modular forms, to the parabolic cohomology groups $\mathrm{H}^1_{\text{ét},!}(\mathrm{Y}_0(N), \mathcal{L}_\ell(k))$.

Following the previous procedure, one would then obtain a 2^d-dimensional ℓ-adic Galois representation of G_F associated with a Hilbert modular form f, instead of what one would expect by the Langlands programme, which is a representation of dimension 2, and not 2^d. The problem is that, working with these Hilbert modular varieties, we have too many complex conjugations at our disposal; and we need to cut them down to a single one. This is done by considering Shimura varieties attached, not to $\boldsymbol{G}$, but to inner forms $\boldsymbol{H}$ of $\boldsymbol{G}$ that are split at a unique infinite place; these are quaternionic Shimura curves. In certain cases one is able to transfer automorphic representations between $\boldsymbol{G}$ and $\boldsymbol{H}$; the method of [Deligne. $\rho_{f,\ell}$] applied to quaternionic Shimura curves allows for the construction of Galois representations attached to Hilbert modular forms whose weights are at least 2 when d is odd [Carayol. $\rho_{f,\ell}$, Théorème A]. This result was then extended to apply even when d is even, using congruences [Taylor. $\rho_{f,\ell}$, Theorem 2] or endoscopic transfer [Blasius–Rogawski, Theorem 1]. The remaining case of partial weight 1 was finally tackled by methods of congruences [Jarvis. $\rho_{f,\ell}$, Theorem 6.1] (the case of parallel weight 1 already having been understood).

Returning to questions of semistability, the application of Theorem I.14.4 to the p-adic Galois representations attached to Hilbert modular forms f prevents the local components $\pi_{f,\mathfrak{p}}$ from being special (i.e. a twist of the Steinberg representation) if the weight of f is too low, where "too low" implies, in particular, that some of the weights are equal to 1. This rules out the study of quaternionic Shimura varieties, and points us towards Hilbert moduli varieties instead.

Wishing to stay within the realm of automorphic representations, the approach to proving a parallel result to Theorem I.14.4 is to study integrality properties of Hecke operators. The primordial observation is given by the conjunction of the following two points:

- if f is Steinberg at p, the Hecke operator U_p acts on the local component $\pi_{f,p}$ of f by multiplication by p^{k-2},
- the Hecke operator U_p acts integrally on the space $\mathrm{M}_k(\Gamma_0(N); \mathbb{C})$ of modular forms of weight k and level $\Gamma_0(N)$, with $p|N$.

These two facts are clearly in contradiction for $k = 1$, wherefore automorphic representations attached to modular forms of weight 1 are nowhere Steinberg. This is to be compared with Theorem I.14.4, which implies that a 2-dimensional filtered (φ, N)-module with Hodge–Tate weights $\{(0, 0)\}$ is necessarily crystalline.

To generalise this observation to the case of Hilbert modular forms, we would like to generalise these two facts. This is a known straightforward task for the former, and relies simply on double cosets decompositions as detailed in section II.3.1, together with the Bernstein–Zelevinsky classification (see section II.2.3). The latter, however, is more subtle. As noted by Hida [Hida. p-adic Hecke algebras, p.297, p.311], when working with Hilbert modular forms of non parallel weight, the Hecke operators incur a certain twist, explained in section II.4.5. Once this twist is accounted for, the appropriate integrality result for $\mathrm{U}_\mathfrak{p}$, proved in section II.5.3 by methods of q-expansions, allows one to deduce Theorem II.6.2, providing a result that matches up with Theorem I.14.4. There is however one important subtlety: whereas Theorem I.14.4, as a result in p-adic Hodge theory, is framed in the setting of local fields, Theorem II.6.2 brings into play the splitting behaviour of p in the totally real field F. This leads to interesting behaviour when the splitting behaviour is complicated, as illustrated in section II.6.2.1.

I

p-adic Hodge theory

1 Hodge theory over $\mathbb{C}$

1.1 The Hodge decomposition theorem

The landmark result of classical Hodge theory is undoubtedly the **Hodge decomposition theorem** for compact Kähler manifolds. This is the decomposition, for a compact Kähler manifold X, of its de Rham cohomology with complex coefficients:

$$\mathrm{H}^n_{\mathrm{dR}}(X, \mathbb{C}) = \bigoplus_{p+q=n} \mathrm{H}^{p,q}(X).$$

Moreover, the individual graded pieces, also called the **Hodge cohomology groups**, satisfy the symmetry property $\overline{\mathrm{H}^{p,q}(X)} = \mathrm{H}^{q,p}(X)$.[3]

This result is of differential geometric nature: we are considering a smooth (C^∞) situation, and working with the sheaves $\mathcal{A}^p_X$ of C^∞ differential p-forms on X, with complex coefficients (we reserve the notation Ω_X for holomorphic forms). The existence of a Kähler metric on X guarantees the existence of the above decomposition of $\mathrm{H}^n_{\mathrm{dR}}(X, \mathbb{C})$, together with the symmetry property for its graded pieces. A typical proof that doesn't stray too far from the statement involves the introduction of the Laplace operator $\Delta := \mathrm{dd}^* + \mathrm{d}^*\mathrm{d}$, where d^* is the adjoint of the exterior derivative d with respect to underlying Riemannian metric on X. Differential forms η with $\Delta\eta = 0$ are called harmonic. One then proves that every differential form can be uniquely written as a sum of a harmonic differential form, an exact differential form (i.e. a differential form in the image of d) and a differential form in the image of d^*.[4] In particular, each $\mathrm{H}^n_{\mathrm{dR}}(X, \mathbb{C})$ has a basis given by harmonic differential n-forms on X.

So far, we've only used the existence of a Riemannian metric on X, and not a Kähler metric. Using the structure of X as a Hermitian complex manifold now, we can split up the Laplace operator into holomorphic and anti-holomorphic components: $\mathrm{d} = \partial + \overline{\partial}$, $\mathrm{d}^* = \partial^* + \overline{\partial}^*$, $\Delta_\partial := \partial\partial^* + \partial^*\partial$, $\Delta_{\overline{\partial}} := \overline{\partial}\overline{\partial}^* + \overline{\partial}^*\overline{\partial}$. Then:

$$\Delta = \mathrm{dd}^* + \mathrm{d}^*\mathrm{d} = (\partial + \overline{\partial})(\partial^* + \overline{\partial}^*) + (\partial^* + \overline{\partial}^*)(\partial + \overline{\partial}) = \Delta_\partial + \Delta_{\overline{\partial}} + (\overline{\partial}\partial^* + \partial^*\overline{\partial}) + (\partial\overline{\partial}^* + \overline{\partial}^*\partial).$$

Now Δ_∂ and $\Delta_{\overline{\partial}}$ are conveniently of bidegree $(0,0)$. Unfortunately, this is not so for the cross terms: $(\overline{\partial}\partial^* + \partial^*\overline{\partial})$ is of bidegree $(-1,1)$, and $(\partial\overline{\partial}^* + \overline{\partial}^*\partial)$ is of bidegree $(1,-1)$. However, supposing these cross terms vanish, so that $\Delta = \Delta_\partial + \Delta_{\overline{\partial}}$ is indeed in bidegree $(0,0)$, we obtain a decomposition: if $\eta = \sum_{p+q=n} \eta_{(p,q)}$ is a harmonic

[3] The notation $\overline{}$ denotes complex conjugation, which is considered with respect to the underlying real subspace, in this case given by $\mathrm{H}^n_{\mathrm{dR}}(X, \mathbb{R}) \otimes_{\mathbb{R}} \mathbb{C} = \mathrm{H}^n_{\mathrm{dR}}(X, \mathbb{C})$ by the universal coefficient theorem.

[4] This is a difficult result, whose usual proofs involve subtle analytic considerations. For instance, one can proceed by constructing two endomorphisms on the space of C^∞ differential forms – the harmonic projector H, and Green's operator G – such that for any differential form η, $\mathrm{H}\eta$ is harmonic, $\mathrm{G}\eta$ is orthogonal to the subspace of harmonic differential forms, and $\eta = \mathrm{H}\eta + \Delta\mathrm{G}\eta$. One method for constructing these operators is to use the heat equation $\partial/\partial t = -\Delta$, as initially done in [Milgram–Rosenbloom]. Writing T_t for the heat transfer evolution operator for time t, T_t smoothes out η to a harmonic asymptotic limit as $t \to +\infty$. This means we can define $\mathrm{H}\eta = \lim_{t\to\infty} \mathrm{T}_t\eta$. As for Green's operator, we can use the formula $\mathrm{G}\eta = \int_0^\infty (\mathrm{T}_t\eta - \mathrm{H}\eta)\mathrm{d}t$, as then $\Delta\mathrm{G}\eta = \int_0^\infty \Delta\mathrm{T}_t\eta\mathrm{d}t = -\int_0^\infty \frac{\partial \mathrm{T}_t\eta}{\partial t}\mathrm{d}t = \eta - H\eta$ as required.

n-form, then so are each of the $\eta_{(p,q)}$. Moreover, if η has real coefficients, then necessarily $\overline{\eta_{(p,q)}} = \eta_{(q,p)}$. Therefore, when the cross terms in the above expansion of the Laplacian vanish, we recover the full statement of the Hodge decomposition theorem. To prove that the cross terms vanish for a Kähler manifold, we use the Kähler identity. Defining first $L\colon \eta \mapsto \eta \wedge \omega$, where ω is the Kähler form of X, and Λ the adjoint of L with respect to the Hermitian metric of X, the fundamental Kähler identity reads $[\Lambda, \partial] = \mathrm{i}\bar{\partial}^*$. A quick calculation shows that this implies that $\Delta = \Delta_\partial + \Delta_{\bar{\partial}}$, and moreover that $\Delta_\partial = \Delta_{\bar{\partial}}$.
This extra identity is useful, in that it allows a more direct definition of the Hodge cohomology groups as **Dolbeault cohomology groups**: $\mathrm{H}^{p,q}_{\mathrm{D}}(X)$ is defined to be the q-th cohomology group of the complex $\mathcal{A}^{p,\bullet}_X$ with differential given by $\bar{\partial}$, where $\mathcal{A}^{p,q}$ is the sheaf of C^∞ differential (p,q)-forms. The formula $\Delta = 2\Delta_{\bar{\partial}}$ then tells us that $\Delta_{\bar{\partial}}$-harmonic forms are the same as Δ-harmonic forms, which identifies Hodge and Dolbeault cohomology groups.

The complexes $\mathcal{A}^\bullet_X$, of C^∞ differential forms, and $\Omega^\bullet_X$, of analytic differential forms, behave somewhat differently: the former is acyclic (as we dispose of partitions of unity in that case), where the latter is not. This entails a certain complication: to correctly compute the cohomology of the complex of analytic sheaves $\Omega^\bullet_X$, we need to replace $\Omega^\bullet_X$ by a quasi-isomorphic complex of acyclic objects before applying the global sections functor.[5] The sheaves $\mathcal{A}^{p,q}$ provide exactly that. They are acyclic, and there is a resolution:

$$\begin{tikzcd}
\Omega^0_X \arrow[r, "\mathrm{d}"] \arrow[d, hook] & \Omega^1_X \arrow[r, "\mathrm{d}"] \arrow[d, hook] & \Omega^2_X \arrow[r, "\mathrm{d}"] \arrow[d, hook] & \Omega^3_X \arrow[r, "\mathrm{d}"] \arrow[d, hook] & \cdots \\
\mathcal{A}^{0,0}_X \arrow[r, "\partial"] \arrow[d, "\bar{\partial}"] & \mathcal{A}^{1,0}_X \arrow[r, "\partial"] \arrow[d, "\bar{\partial}"] & \mathcal{A}^{2,0}_X \arrow[r, "\partial"] \arrow[d, "\bar{\partial}"] & \mathcal{A}^{3,0}_X \arrow[r, "\partial"] \arrow[d, "\bar{\partial}"] & \cdots \\
\mathcal{A}^{0,1}_X \arrow[r, "\partial"] \arrow[d, "\bar{\partial}"] & \mathcal{A}^{1,1}_X \arrow[r, "\partial"] \arrow[d, "\bar{\partial}"] & \mathcal{A}^{2,1}_X \arrow[r, "\partial"] \arrow[d, "\bar{\partial}"] & \mathcal{A}^{3,1}_X \arrow[r, "\partial"] \arrow[d, "\bar{\partial}"] & \cdots \\
\mathcal{A}^{0,2}_X \arrow[r, "\partial"] \arrow[d, "\bar{\partial}"] & \mathcal{A}^{1,2}_X \arrow[r, "\partial"] \arrow[d, "\bar{\partial}"] & \mathcal{A}^{2,2}_X \arrow[r, "\partial"] \arrow[d, "\bar{\partial}"] & \mathcal{A}^{3,2}_X \arrow[r, "\partial"] \arrow[d, "\bar{\partial}"] & \cdots \\
\mathcal{A}^{0,3}_X \arrow[r, "\partial"] \arrow[d, "\bar{\partial}"] & \mathcal{A}^{1,3}_X \arrow[r, "\partial"] \arrow[d, "\bar{\partial}"] & \mathcal{A}^{2,3}_X \arrow[r, "\partial"] \arrow[d, "\bar{\partial}"] & \mathcal{A}^{3,3}_X \arrow[r, "\partial"] \arrow[d, "\bar{\partial}"] & \cdots \\
\vdots & \vdots & \vdots & \vdots &
\end{tikzcd}$$

Can we use this data to recover the Hodge decomposition theorem? Deligne's observation [Deligne. Hodge II, Proposition 2.1.9] is that we should focus on filtrations instead of direct sum decompositions, with his introduction of the notion of **opposed filtrations**. Instead of considering the individual $\mathrm{H}^{p,q}(X)$ and the decomposition $\bigoplus_{p+q=n} \mathrm{H}^{p,q}(X)$ of $\mathrm{H}^n_{\mathrm{dR}}(X, \mathbb{C})$, we should focus rather on the purely holomorphic object which is the **Hodge filtration** $F^\bullet \mathrm{H}^n_{\mathrm{dR}}(X, \mathbb{C})$ given by

$$F^p \mathrm{H}^n_{\mathrm{dR}}(X, \mathbb{C}) := \bigoplus_{\substack{s \geqslant p \\ s+q=n}} \mathrm{H}^{s,q}(X),$$

and the conjugate filtration $\bar{F}^\bullet$ defined analogously. This is more tractable algebraically: there is no mixing of holomorphic and conjugate-holomorphic parts.[6] We will see, in the next section, how an algebraic definition of this filtration is possible, making use of double complexes like the one above.
The symmetry condition $\overline{\mathrm{H}^{p,q}(X)} = \mathrm{H}^{q,p}(X)$ then corresponds, in terms of filtrations, to the condition that $F^\bullet$ and $\bar{F}^\bullet$ on $H = \mathrm{H}^n_{\mathrm{dR}}(X, \mathbb{C})$ be *n-opposed*: $\mathrm{gr}^p_F \mathrm{gr}^q_{\bar{F}} H = 0$ whenever $p + q \neq n$. The Hodge decomposition can

[5] This procedure is necessary to calculate the correct homotopical information. The problem is that the global sections functor does not preserve homotopy, which is rectified by only ever applying it to injective objects (this is the notion of cofibrant replacement in homotopical algebra).
[6] In other words, the filtration $F^\bullet$ varies holomorphically in families, whereas the Hodge decomposition does not. See section 1.2.1.

then be obtained from the filtrations $F^\bullet$ and $\bar{F}^\bullet$ by defining $H^{p,q} = F^p H \cap \bar{F}^q H$ whenever $p + q = n$ (and 0 otherwise).

1.2 Algebraic de Rham cohomology

In the algebraic setting, we instead consider the case of a smooth algebraic variety X over a field K, and the sheaves of *algebraic* Kähler differentials $\Omega^i_{X/K}$ of X. The sheaves are not acyclic, so again we will need to take a resolution. Another difficulty also crops up: whereas in the C^∞ and holomorphic situations one disposes of a Poincaré lemma, this is no longer the case algebraically (for instance, $1/z$ does not, even locally, admit an algebraic primitive). This means that the analytic de Rham complex is not even exact!

We need to use tools from the derived toolbox to compute cohomology using resolutions. The standard procedure involves the notion of **hypercohomology**. To replace our original complex with a quasi-isomorphic injective object (in the category of chain complexes), we can resolve each degree individually, vertically. The best way to proceed was figured out by Cartan and Eilenberg.
Our situation: we have an abelian category $\boldsymbol{A}$ (here the category of sheaves of K-vector spaces on X) with enough injectives, and a chain complex $(\boldsymbol{C}^\bullet, d)$ of objects of $\boldsymbol{A}$. We are interested in computing cohomology over X, i.e. the right derived functors of the covariant left-exact global sections functor $\Gamma\colon \mathfrak{F} \mapsto \mathfrak{F}(X)$.[7] The fundamental result that will allow us to compute cohomology of complexes is then [Cartan–Eilenberg, Chapter XVII, Proposition 1.2]:

Theorem 1.1 (Cartan–Eilenberg) — There exists a double complex $I = (I^{\bullet,\bullet}, \partial, \bar{\partial})$ of injective objects of $\boldsymbol{A}$, with a diagram

$$\begin{tikzcd}
C^0 \arrow[r, "d"] \arrow[d, hook] & C^1 \arrow[r, "d"] \arrow[d, hook] & C^2 \arrow[r, "d"] \arrow[d, hook] & C^3 \arrow[r, "d"] \arrow[d, hook] & \cdots \\
I^{0,0} \arrow[r, "\partial"] \arrow[d, "\bar\partial"] & I^{1,0} \arrow[r, "\partial"] \arrow[d, "\bar\partial"] & I^{2,0} \arrow[r, "\partial"] \arrow[d, "\bar\partial"] & I^{3,0} \arrow[r, "\partial"] \arrow[d, "\bar\partial"] & \cdots \\
I^{0,1} \arrow[r, "\partial"] \arrow[d, "\bar\partial"] & I^{1,1} \arrow[r, "\partial"] \arrow[d, "\bar\partial"] & I^{2,1} \arrow[r, "\partial"] \arrow[d, "\bar\partial"] & I^{3,1} \arrow[r, "\partial"] \arrow[d, "\bar\partial"] & \cdots \\
I^{0,2} \arrow[r, "\partial"] \arrow[d, "\bar\partial"] & I^{1,2} \arrow[r, "\partial"] \arrow[d, "\bar\partial"] & I^{2,2} \arrow[r, "\partial"] \arrow[d, "\bar\partial"] & I^{3,2} \arrow[r, "\partial"] \arrow[d, "\bar\partial"] & \cdots \\
I^{0,3} \arrow[r, "\partial"] \arrow[d, "\bar\partial"] & I^{1,3} \arrow[r, "\partial"] \arrow[d, "\bar\partial"] & I^{2,3} \arrow[r, "\partial"] \arrow[d, "\bar\partial"] & I^{3,3} \arrow[r, "\partial"] \arrow[d, "\bar\partial"] & \cdots \\
\vdots & \vdots & \vdots & \vdots &
\end{tikzcd}$$

such that the columns are individually injective resolutions of their top components, and that moreover the horizontal kernel, image, and cohomology complexes are also injective resolutions of the respective kernels, images and cohomologies of the top row.[8] ◇□

In such a situation, the total complex $T^\bullet = \mathrm{Tot}^\bullet I$ of I is then an injective complex which is quasi-isomorphic to $\boldsymbol{C}^\bullet$. The correct object to consider, instead of the naive cohomology groups $\mathrm{H}^i(C^j)$, is the hypercohomology

$$\mathbf{H}^n(\boldsymbol{C}^\bullet) = \mathbf{R}^n\Gamma(\boldsymbol{C}^\bullet) = \mathrm{H}^\bullet(X, T^\bullet).$$

In particular, this allows us to define the **algebraic de Rham cohomology** of X as the hypercohomology of the complex $\Omega^\bullet_{X/K}$:

$$\mathrm{H}^n_{\mathrm{dR}}(X/K) := \mathbf{H}^n(X, \Omega^\bullet_{X/K}).$$

[7] More generally, the following applies to any covariant left-exact additive functor $\Gamma\colon \boldsymbol{A} \to \boldsymbol{B}$ between abelian categories. Changing covariant to contravariant, or left-exact to right exact, means we instead have to consider projective objects. Injective objects will also work for a contravariant right-exact functor.

[8] Such a double complex is called a Cartan–Eilenberg resolution of $\boldsymbol{C}^\bullet$.

Now, how do we get our filtrations? Ideally we would want to obtain two n-opposed filtrations on $\mathrm{H}^n_{\mathrm{dR}}(X/K)$ that generalise the Hodge filtration in the complex case. Unfortunately, this isn't always possible.
To understand this, we first define two filtrations on the total complex $S^\bullet$ of $A^{\bullet,\bullet} = \Gamma I^{\bullet,\bullet}$, where I is a Cartan–Eilenberg resolution of $\Omega^\bullet_{X/K}$. These are the horizontal filtration

$$(F^p S)^n := \bigoplus_{\substack{s+q=n \\ s \geqslant p}} A^{s,q},$$

and the vertical filtration

$$(G^q S)^n := \bigoplus_{\substack{p+t=n \\ t \geqslant q}} A^{p,t}.$$

These filtrations then have associated spectral sequences with initial pages

$$'\mathrm{E}_1^{p,q} = \mathrm{H}^q(X, \Omega^p_{X/K}),$$
$$''\mathrm{E}_1^{p,q} = \mathrm{H}^p(X, \Omega^q_{X/K}), \qquad ''\mathrm{E}_2^{p,q} = \mathrm{H}^p\left(X, \mathcal{H}^q(\Omega^\bullet_{X/K})\right),$$ [9]

which both converge to the hypercohomology $\mathbf{H}^n(X, \Omega^\bullet_{X/K})$, i.e. the algebraic de Rham cohomology of X. These are the **first and second hypercohomology spectral sequences**. In this situation (i.e. as opposed to a more general functor than the global sections functor), we also call the first spectral sequence the **Hodge-to-de Rham spectral sequence**, as it links the Hodge cohomology groups $\mathrm{H}^q(X, \Omega^p_{X/K})$ with the de Rham cohomology groups.
From the construction of the spectral sequence associated with a filtration, we know that the abutment of each spectral sequence inherits a filtration, from F (respectively G). In the first case, the filtration is also obtained as the image of the **trivial filtration** $\sigma^{\geqslant p}$ of $\Omega^\bullet_{X/K}$, i.e.

$$(\sigma^{\geqslant p} C^\bullet)^n := \begin{cases} 0 & \text{if } n < p, \\ C^n & \text{if } n \geqslant p, \end{cases}$$

whereas for the second case, the abutment filtration comes from the **canonical filtration** $\tau_{\leqslant p}$, given by:

$$(\tau_{\leqslant p} C^\bullet)^n := \begin{cases} C^n & \text{if } n < p, \\ \ker(\mathrm{d}) & \text{if } n = p, \\ 0 & \text{if } n \geqslant p. \end{cases}$$

The correspondence between the spectral sequence of the filtered complex $(\Omega^\bullet_{X/K}, \tau_\leqslant)$ and the second hypercohomology spectral sequence is given by the renumbering $_\tau\mathrm{E}_r^{p,q} = {''\mathrm{E}}_{r+1}^{2p+q,-p}$.

After all this, we have indeed acquired two filtrations on $\mathrm{H}^n_{\mathrm{dR}}(X/K)$. We have, however, two problems.

- Consider the first hypercohomology spectral sequence. To get a decomposition of the algebraic de Rham cohomology into Hodge cohomology groups, $\mathrm{H}^n_{\mathrm{dR}}(X/K) = \bigoplus_{p+q=n} \mathrm{H}^q(X, \Omega^p_{X/K})$, we can't afford for there to be any non-zero differentials on the $'\mathrm{E}_1$ page. If the spectral sequence doesn't degenerate at $'\mathrm{E}_1$, some cocyles will be killed off.

- We also don't know in general whether the two filtrations obtained by the two hypercohomology spectral sequences will be opposed.

To understand this situation, it is helpful to get a grounding on the structure of double complexes over K-vector spaces. It is known that in the category of bounded double complexes of K-vector spaces, every object is a direct sum of indecomposable objects, which are of the following three forms:
dots, that are non-zero in a single bidegree: $[K]$,

[9] Here $\mathcal{H}^i(\mathfrak{F})$ denotes the i-th cohomology sheaf of $\mathfrak{F}$ on X.

squares:

$$\left[\begin{array}{ccc} K & \xrightarrow{\sim} & K \\ \wr\big\downarrow & & \big\downarrow\wr \\ K & \xrightarrow{\sim} & K \end{array}\right],$$

and **zigzags**:

$$\left[\begin{array}{ccc} K & \xrightarrow{\sim} & K \end{array}\right], \qquad \left[\begin{array}{c} K \\ \big\downarrow\wr \\ K \end{array}\right], \qquad \left[\begin{array}{ccc} K & \xrightarrow{\sim} & K \\ \big\downarrow\wr & & \\ K & & \end{array}\right], \qquad \left[\begin{array}{ccccc} & & K & \xrightarrow{\sim} & K \\ & & \big\downarrow\wr & & \\ K & \xrightarrow{\sim} & K & & \end{array}\right], \qquad \cdots$$

The length of a zigzag is defined to be the number of non-zero terms, one more than the number of non-zero arrows. By definition then, the length of a zigzag is always at least 2.

Consider now a bounded double complex $A = (A^{\bullet,\bullet}, \partial, \bar{\partial})$, with the horizontal and vertical filtrations F,G on its total complex $T = (T^\bullet, d)$ defined as above. We have the following elementary observations:

- The filtrations on $\mathrm{H}^n(T^\bullet)$ induced by F and G are n-opposite if and only if A has no odd-length zigzags in its decomposition.
- The differential d is strictly compatible with the filtration F (respectively G) if and only if A has no even-length zigzags that start (and end) with vertical (respectively horizontal) arrows in its decomposition. We recall that the strict compatibility of d with the filtration F is the condition that $d(F^pT) = (F^pT) \cap d(A)$; it is equivalent to the degeneration of the spectral sequence attached to the filtration F at the E_1 page (and ditto for G) [Deligne. Hodge II, Proposition I.3.2].

This then tells us what we can expect in general. We will always have two filtrations F and G on $\mathrm{H}^n_{\mathrm{dR}}(X/K)$, but they will only have the expected graded pieces if the associated hypercohomology spectral sequences degenerate at E_1, which happens if and only if the double complex $\Gamma(I^{\bullet,\bullet})$, where I is a Cartan–Eilenberg resolution of $\Omega^\bullet$, has no even-length zigzags. On the other hand, the filtrations are opposed if and only if $\Gamma(I^{\bullet,\bullet})$ has no odd-length zigzags.

1.2.1 Relative algebraic de Rham cohomology

The upshot of viewing the Hodge filtration through the lens of the hypercohomology spectral sequences is that it provides a direct algebraic definition, as opposed to the definition through harmonic forms which has the appearance of depending on the specific Kähler structure chosen. In particular, we have a way of seeing that the Hodge filtration varies algebraically in families.

Indeed, consider now a smooth morphism $f\colon X \to S$ over a field K. Proceeding by analogy with the definition of algebraic de Rham cohomology in the absolute situation, we define the **relative de Rham cohomology** of f as $\mathscr{H}_{\mathrm{dR}}(X/S) := \mathbf{R}f_*\Omega^\bullet_{X/S}$. Subsequently, $\mathbf{R}^q f_*\Omega^\bullet_{X/S}$ inherits a Hodge filtration from the trivial filtration on $\Omega^\bullet_{X/S}$:

$$F^p\mathbf{R}^q f_*\Omega^\bullet_{X/S} = \mathrm{im}\left(\mathbf{R}^q f_*\left(\sigma^{\geqslant p}\Omega^\bullet_{X/S}\right) \longrightarrow \mathbf{R}^q f_*\Omega^\bullet_{X/S}\right).$$

Associated to this filtration is the relative Hodge-to-de Rham spectral sequence:

$${}_F\mathrm{E}_1^{p,q} = \mathrm{R}^q f_*\Omega^p_{X/S} \Rightarrow \mathbf{R}^n f_*\Omega^\bullet_{X/S}.$$

To study the variation of this filtration, we define the **Koszul filtration** $\mathrm{K}^\bullet$ on $\Omega^\bullet_{X/K}$ as

$$\mathrm{K}^i\Omega^\bullet_{X/K} := \mathrm{im}\left(f^*\Omega^i_{S/K} \otimes_{\mathcal{O}_X} \Omega^\bullet_{X/K}[-i] \xrightarrow{\wedge} \Omega^\bullet_{X/K}\right).$$

The interest in this filtration comes from the study of the associated spectral sequence. Indeed, its first page is given by

$$ {}_{\mathrm{K}}\mathrm{E}_1^{p,q} = \mathbf{R}^{p+q} f_* \left(\mathrm{gr}^p_{\mathrm{K}} \Omega^\bullet_{X/K} \right) = \mathbf{R}^q f_* \left(f^* \Omega^p_{S/K} \otimes_{\mathcal{O}_X} \Omega^\bullet_{X/S} \right) = \Omega^p_{S/K} \otimes_{\mathcal{O}_S} \mathbf{R}^q f_* \Omega^\bullet_{X/S}, $$

where the above equalities follow from the exact sequence

$$ 0 \to f^* \Omega^\bullet_{S/K} \to \Omega^\bullet_{X/K} \to \Omega^\bullet_{X/S} \to 0, $$

together with the projection formula.

Now, for fixed q, the differentials $d_1^{p,q}$ of the ${}_{\mathrm{K}}\mathrm{E}_1^{p,q}$ page fit together to give a complex

$$ 0 \longrightarrow \mathcal{H}^q_{\mathrm{dR}}(X/S) \xrightarrow{d_1^{0,q}} \Omega^1_{S/K} \otimes_{\mathcal{O}_S} \mathcal{H}_{\mathrm{dR}}(X/S) \xrightarrow{d_1^{1,q}} \Omega^2_{S/K} \otimes_{\mathcal{O}_S} \mathcal{H}_{\mathrm{dR}}(X/S) \xrightarrow{d_1^{2,q}} \cdots . $$

Of particular interest is the morphism $d_1^{0,q} \colon \mathcal{H}^q_{\mathrm{dR}}(X/S) \longrightarrow \Omega^1_{S/K} \otimes_{\mathcal{O}_S} \mathcal{H}_{\mathrm{dR}}(X/S)$, which certainly has the expected appearance of a Koszul connection on the sheaf $\mathcal{H}^q_{\mathrm{dR}}(X/S)$. For $d_1^{0,q}$ to be a bona-fide Koszul connection, by definition we just need to check it satisfies the Leibniz rule, i.e. that for local sections α of $\Omega^i_{S/K}$ and s of $\mathcal{H}^q_{\mathrm{dR}}(X/S)$, the formula

$$ d_1^{i,q}(\alpha \cdot s) = \mathrm{d}\alpha + (-1)^i \alpha \cdot d_1^{0,q} s $$

is satisfied. This formula in fact follows from the product structure on the spectral sequence [Katz–Oda, §2], inherited from the compatibility of $\mathrm{K}^\bullet$ with wedge products. This then means that $d_1^{0,q}$ is a Koszul connection on $\mathcal{H}^q_{\mathrm{dR}}(X/S)$; moreover the other differentials $d_1^{i,q}$ are uniquely determined by $d_1^{0,q}$ by the Leibniz rule. In addition, because the ${}_{\mathrm{K}}\mathrm{E}_1^{p,q}$ page is a complex under the differentials $d_1^{p,q}$, we have that $d_1^{1,q} \circ d_1^{0,q} = 0$, so that $d_1^{0,q}$ is in fact a *flat* connection. It is called the **Gauss–Manin connection**, and is usually denoted ∇.

This is a somewhat roundabout definition, but the Gauss–Manin connection is the natural connection to impose on the relative algebraic de Rham cohomology of a smooth morphism. In particular, if we are working over $\mathbb{C}$, we can use the relative version of the de Rham theorem (which compares de Rham cohomology and singular cohomology). In that framework, on top of working with $\mathbf{R} f_* \Omega^\bullet_{X/S}$, we also have recourse to the locally constant sheaf $\mathrm{R}^q f_* \mathbb{Z}$, with $(\mathrm{R}^q f_* \mathbb{Z})_s = \mathrm{H}^q(X_s, \mathbb{Z})$. The crucial check is to make sure that the sheaf of ∇-horizontal sections of $\mathbf{R}^q f_* \Omega^\bullet_{X/S}$ is in fact naturally isomorphic to $\mathrm{R}^q f_* \mathbb{C}$ [Deligne. ∇, Proposition I.2.28]. This shows that the Gauss–Manin connection is indeed the correct object to consider; it genuinely accounts for the variation between the fibres of the locally free sheaves $\mathbf{R}^q f_* \Omega^\bullet_{X/S}$.

We have the following fundamental observation about the interaction of ∇ with the Hodge filtration.

Theorem 1.2 (Griffiths transversality) — The Gauss–Manin connection ∇ satisfies the formula

$$ \nabla F^p \mathcal{H}_{\mathrm{dR}}(X/S) \subseteq \Omega^1_{S/K} \otimes_{\mathcal{O}_S} F^{p-1} \mathcal{H}_{\mathrm{dR}}(X/S). \qquad \diamond $$

Proof: The differential $d_1^{0,q}$ can be described as a connecting homomorphism. We have an exact sequence

$$ 0 \longrightarrow \mathrm{K}^1 \Omega^\bullet_{X/K} / \mathrm{K}^2 \Omega^\bullet_{X/K} \longrightarrow \mathrm{K}^0 \Omega^\bullet_{X/K} / \mathrm{K}^2 \Omega^\bullet_{X/K} \longrightarrow \mathrm{K}^0 \Omega^\bullet_{X/K} \longrightarrow 0. $$

We can rewrite the outermost terms more explicitly:

$$ 0 \longrightarrow f^* \Omega^1_{S/K} \otimes_{\mathcal{O}_X} \Omega^\bullet_{X/S}[-1] \longrightarrow \mathrm{K}^0 \Omega^\bullet_{X/K} / \mathrm{K}^2 \Omega^\bullet_{X/K} \longrightarrow \Omega^\bullet_{X/S} \longrightarrow 0. $$

Then $d_1^{0,q}$ is the associated connecting homomorphism arising from the application of $\mathbf{R}^q f_*$, as a result of the construction of the differentials on the first page of the spectral sequence attached to a filtered complex [Katz–Oda, §3, *Reduction.*].

If we apply the truncation functor $\sigma^{\geqslant p}$ before $\mathbf{R} f_*$, we get:

$$ 0 \longrightarrow f^* \Omega^1_{S/K} \otimes_{\mathcal{O}_X} \sigma^{\geqslant p-1} \Omega^\bullet_{X/S}[-1] \longrightarrow \sigma^{\geqslant p} \left(\mathrm{K}^0 \Omega^\bullet_{X/K} / \mathrm{K}^2 \Omega^\bullet_{X/K} \right) \longrightarrow \sigma^{\geqslant p} \Omega^\bullet_{X/S} \longrightarrow 0. $$

The corresponding connecting homomorphism is then

$$d_1^{0,q} \colon F^p \mathbf{R}^q f_* \Omega^\bullet_{X/S} \longrightarrow F^{p-1} \mathbf{R}^q f_* \left(f^* \Omega^1_{S/K} \otimes_{\mathbb{O}_X} \Omega^\bullet_{X/S} \right) = \Omega^1_{S/K} \otimes_{\mathbb{O}_S} F^{p-1} \mathbf{R}^q f_* \Omega^\bullet_{X/S}. \qquad \square$$

This result tells us how the Hodge filtration behaves in families: starting with a section in $F^p \mathcal{H}_{\mathrm{dR}}(X/S)$ and horizontally transporting it along some vector field through the connection ∇, we are guaranteed to land in $F^{p-1} \mathcal{H}_{\mathrm{dR}}(X/S)$. In particular, the Hodge filtration varies algebraically in families.

1.3 Hodge structures

It is useful to introduce some axiomatics capturing the situation of the Hodge decomposition theorem. Given a finite dimensional $\mathbb{R}$-vector space V, a **Hodge structure** on V consists of a bigrading on the complexification $V_{\mathbb{C}} = V \otimes_{\mathbb{R}} \mathbb{C}$ of V, say $V_{\mathbb{C}} = \bigoplus_{p,q} V^{p,q}$, satisfying $\overline{V^{p,q}} = V^{q,p}$. The weight grading of such an object is given by $V^{(k)} = \bigoplus_{p+q=k} V^{p,q}$, and we say the Hodge structure is of weight n if $V = V^{(n)}$, i.e. $V^{p,q} = 0$ for $p + q \neq n$. The type of a Hodge structure is the set of pairs (p, q) such that $V^{p,q} \neq 0$. The Hodge decomposition theorem can then be recast as saying that $\mathrm{H}^n_{\mathrm{dR}}(X, \mathbb{R})$ comes naturally equipped with Hodge structure of weight n.

There are several reformulations possible. As alluded to previously, a Hodge structure of weight n on V is the same as the data of a finite filtration $F^\bullet$ on $V_{\mathbb{C}}$ which is n-opposed to its complex conjugate filtration $\overline{F}^\bullet$ [Deligne. Hodge II, Proposition 2.1.9]. Another angle is to consider the category of Hodge structures (with the obvious notion of morphism) as a category of representations. The category of graded vector spaces is the category of representations of $\mathbb{G}_{\mathrm{m}}$, with the i-th graded piece corresponding to the isotypic component under the i-th power character of $\mathbb{G}_{\mathrm{m}}$; similarly the category of bigraded vector spaces is the category of representations of $\mathbb{G}_{\mathrm{m}} \times \mathbb{G}_{\mathrm{m}}$. In our situation, the relevant group is the real algebraic group called the **Deligne torus** $\mathbb{S} = \mathrm{Res}^{\mathbb{C}}_{\mathbb{R}} \mathbb{G}_{\mathrm{m},\mathbb{C}}$, whose complexification $\mathbb{S}_{\mathbb{C}}$ is two copies of $\mathbb{G}_{\mathrm{m},\mathbb{C}}$ interchanged by complex conjugation. Thus a third possible description is that a Hodge structure is a representation of $\mathbb{S}$ on a finite dimensional real vector space. We can also define the notion of Hodge structures on, say, R-modules for other rings R than $\mathbb{R}$ [Deligne. Hodge II, 2.1.12] : given a Noetherian subring $R \subseteq \mathbb{R}$ and a projective R-module M, a Hodge structure on M consists of a (real) Hodge structure on $M_{\mathbb{R}} = M \otimes_R \mathbb{R}$ such that its weight grading is defined over $\mathrm{Frac}(R)$. This is useful as we can refine the $\mathbb{R}$-Hodge structure on $\mathrm{H}^n_{\mathrm{dR}}(X, \mathbb{R})$ to a $\mathbb{Z}$-Hodge structure: the singular cohomology with integer coefficients $\mathrm{H}^n_{\mathrm{sing}}(X, \mathbb{Z})$ yields a lattice $\mathrm{H}^n_{\mathrm{sing}}(X, \mathbb{Z}) \subseteq \mathrm{H}^n_{\mathrm{dR}}(X, \mathbb{R})$, through the de Rham isomorphism between singular and de Rham cohomology.

One subtle but crucial consideration is that of rank one Hodge structures. If X is a connected compact Kähler manifold of (complex) dimension d, we can integrate top forms

$$\int_X \colon \mathrm{H}^{2d}_{\mathrm{dR}}(X, \mathbb{R}) \to \mathbb{C}.$$

The important observation is that, to do so, we need to pick an orientation on X, which reduces to the choice of i, a square root of -1, also corresponding to a generator γ of $\mathrm{H}_1^{\mathrm{sing}}(\mathbb{C}^\times, \mathbb{Z})$. The dual element in $\mathrm{H}^1_{\mathrm{dR}}(\mathbb{C}^\times, \mathbb{C})$ to such a generator is the differential form $\frac{1}{2\pi\mathrm{i}} \frac{\mathrm{d}z}{z}$, because of Euler's formula

$$\int_\gamma \frac{\mathrm{d}z}{z} = 2\pi\mathrm{i}.$$

The upshot is that, even though $\mathrm{H}^1_{\mathrm{sing}}(\mathbb{C}^\times, \mathbb{Z})$ is not naturally isomorphic to $\mathbb{Z}$, it is naturally isomorphic to the rank one $\mathbb{Z}$-sublattice of $\mathbb{C}$ given by $(2\pi\mathrm{i})^{-1}\mathbb{Z} \subset \mathbb{C}$, independently of choices of orientation. Similarly, integration canonically identifies $\mathrm{H}^{2d}_{\mathrm{sing}}(X, \mathbb{Z}) \subseteq \mathrm{H}^{2d}_{\mathrm{dR}}(X, \mathbb{C})$ with $(2\pi\mathrm{i})^{-d}\mathbb{Z} \subset \mathbb{C}$, independently of any choices of orientations.

This motivates the definition of **Tate's Hodge structures**: $\mathbb{Z}(i)$ is the $\mathbb{Z}$-Hodge structure of rank 1 and type $\{(-i, -i)\}$ (and thus weight $-2i$) given by the $\mathbb{Z}$-module $(2\pi\mathrm{i})^i\mathbb{Z} \subset \mathbb{C}$. We also define $R(i) = \mathbb{Z}(i) \otimes_{\mathbb{Z}} R$. The integration operator above can then be seen as providing a canonical isomorphism of $\mathbb{Z}$-Hodge structures

$\mathrm{H}^{2d}_{\mathrm{sing}}(X,\mathbb{Z}) \cong \mathbb{Z}(-d)$, which no longer depends on any choices of orientations.
The distinction between $\mathbb{Z}$ and $\mathbb{Z}(i)$ takes on a particularly important meaning once one tries to compare the complex theory with other algebraic theories.[10]

1.4 Integration on varieties over $\mathbb{C}$: the complex comparison theorem

The point of inspiration for p-adic Hodge theory is the integration pairing for a compact complex manifold X:

$$\mathrm{H}^n_{\mathrm{dR}}(X,\mathbb{C}) \times \mathrm{H}^{\mathrm{sing}}_n(X,\mathbb{C}) \to \mathbb{C},$$

which is perfect by de Rham's theorem.
The most famous instance of this is undoubtedly Euler's formula

$$\int_\gamma \frac{\mathrm{d}z}{z} = 2\pi\mathrm{i},$$

which expresses $2\pi\mathrm{i}$ as a period: the line integral of $\frac{\mathrm{d}z}{z}$ around the circle $\mathrm{S}^1 \subset \mathbb{C}^\times$.[11]

This shows that the integration pairing fundamentally requires the use of complex numbers: ${}^{\mathrm{d}z}/_{z}$ has integer coefficients and is a basis for $\mathrm{H}^1_{\mathrm{dR}}(\mathbb{C}^\times,\mathbb{C})$, γ is a basis for $\mathrm{H}^{\mathrm{sing}}_1(\mathbb{C}^\times,\mathbb{Z})$, yet $\int_\gamma {}^{\mathrm{d}z}/_{z} = 2\pi\mathrm{i}$, which is a transcendental imaginary number!
In the case where X is instead a smooth projective variety over a subfield K of $\mathbb{C}$, the integration pairing would then look like:

$$\mathrm{H}^n_{\mathrm{dR}}(X/K) \times \mathrm{H}^{\mathrm{sing}}_n(X(\mathbb{C}),\mathbb{Z}) \to \mathbb{C},$$

which corresponds to an isomorphism between $\mathrm{H}^n_{\mathrm{dR}}(X/K)$ and $\mathrm{H}^n_{\mathrm{sing}}(X(\mathbb{C}),\mathbb{Z})$ *after they have both been tensored up to* $\mathbb{C}$.
There are a surprising number of important properties of the complex numbers that come together to make this result work: in particular, $\mathbb{C}$ is both complete and algebraically closed. Perhaps less appreciated is the suitability of the complex topology for manipulating differentials.

2 p-adic integration?

Suppose we wanted a similar integration pairing p-adically. Start with a smooth projective variety X/K, with K a finite extension of $\mathbb{Q}_p$. To talk about integration of one-forms over paths on X, we need a p-adic analytic topology, so we should start at least with $\mathbb{C}_p$, the completion of the algebraic closure of K, which is itself algebraically closed by Krasner's lemma. However the p-adic topology on $\mathbb{C}_p$ is *not* fine enough for the purposes of integrating differentials. Indeed, the problem is that in going from K to $\overline{K}$, we have picked up a lot of ramification, too much in fact for the p-adic topology to cope with. If we consider the universal derivation $\mathrm{d}\colon \mathcal{O}_{\overline{K}} \to \Omega^1_{\mathcal{O}_{\overline{K}}/\mathcal{O}_K}$, we will see that $\ker(\mathrm{d})$ does not contain any p-adic neighbourhood of 0, $p^n\mathcal{O}_{\overline{K}}$, in $\mathcal{O}_{\overline{K}}$. This means that any notion of a p-adic integral of differential 1-forms, if it takes values in $\mathbb{C}_p$, is susceptible to confound quantities whose difference is p-adically indistinguishable from 0, but which would have been separated from 0 by some open set of the form $p^n\ker(\mathrm{d})$ in a completion of $\overline{K}$ with respect to the finer topology using $p^n\ker(\mathrm{d})$ as neighbourhoods of 0 in $\overline{K}$.
Indeed, we can try to compute the same integral as before: $\int_\gamma \frac{\mathrm{d}z}{z}$, with γ describing a closed loop around 0 on the rigid analytic space associated with $\mathbb{G}_{\mathrm{m},K}$. To avoid a long discussion on the correct definitions of rigid analytic spaces, and integrals on such spaces, we note that any such definition would necessarily satisfy:

$$\int_\gamma \frac{\mathrm{d}z}{z} = p^n \int_1^{\varepsilon_n} \frac{\mathrm{d}z}{z} = p^n \log(\varepsilon_n),$$

[10] One such example is ℓ-adic étale cohomology, where the distinction becomes about remembering the Galois action: in that context, $\mathbb{Q}_\ell(i)$ is a one-dimensional $\mathbb{Q}_\ell$-vector space which has a non-trivial Galois action, given by χ_ℓ^i, where χ_ℓ is the ℓ-adic cyclotomic character.
[11] As above, the orientation γ on S^1 is implicitly determined by the choice of i.

where log denotes any p-adic logarithm defined on $\mathbb{C}_p^\times$, and $\varepsilon = (\varepsilon_n)_n$ is a p-adic orientation, i.e. a compatible choice of p-power roots of unity: $\varepsilon_0 = 1$, ε_1 a primitive p-th root of unity, $\varepsilon_n^p = \varepsilon_{n-1}$. Hence, as $\{\log(\varepsilon_n) : n \in \mathbb{N}\}$ is bounded in $\mathbb{C}_p$, we have that:

$$\int_\gamma \frac{\mathrm{d}z}{z} = \lim_{n\to\infty} p^n \log(\varepsilon_n) = 0.$$

We see then that in the p-adic topology, $(p^n \log(\varepsilon_n))_n$ tends to 0, however this quantity does not tend to 0 if one refines the topology of $\mathbb{C}_p$ with subsets corresponding to $\ker(\mathrm{d})$, as discussed above. This will become apparent once we define the correct completion of $\overline{K}$, which is usually denoted $\mathbf{B}_{\mathrm{dR}}^+$.

3 Three views of $\mathbf{B}_{\mathrm{dR}}^+$

In the end, we see that to define a good notion of p-adic integration, we require a sufficiently fine topology on the ring of coefficients to allow differential calculus.

A direct approach is possible: we start by giving $\ker(\mathrm{d}) \subset \mathcal{O}_{\overline{K}}$ the topology generated by neighbourhoods of 0 of the form $p^n \ker(\mathrm{d})$, and completing. This yields the universal p-adic first-order infinitesimal $\mathcal{O}_K$-thickening of $\mathcal{O}_{\mathbb{C}_p}$ [Fontaine. $\mathbf{B}_{\mathrm{dR}}$, Proposition 1.4.3]. This object, as opposed to $\mathcal{O}_{\mathbb{C}_p}$, has a topology fine enough to perform manipulations with algebraic differential 1-forms over K. If we want higher order infinitesimal neighbourhoods, following [Colmez. $\mathbf{B}_{\mathrm{dR}}^+$, §3], we can inductively define:

- $\mathcal{O}_{\overline{K}}^{(0)} := \mathcal{O}_{\overline{K}}$,
- $\Omega^{(k)} := \mathcal{O}_{\overline{K}} \otimes_{\mathcal{O}_{\overline{K}}^{(k-1)}} \Omega^1_{\mathcal{O}_{\overline{K}}^{(k-1)}/\mathcal{O}_K}$, each a torsion $\mathcal{O}_{\overline{K}}$-module,
- $\mathcal{O}_{\overline{K}}^{(k)} := \ker\left(\mathrm{d}\colon \mathcal{O}_{\overline{K}}^{(k-1)} \to \Omega^{(k)}\right)$, a decreasing sequence of subrings of $\mathcal{O}_{\overline{K}}$.

Now, if we complete $\mathcal{O}_{\overline{K}}$ with respect to the topology generated by the open neighbourhoods of 0 given by $U_{n,k} := p^n \mathcal{O}_{\overline{K}}^{(k)}$, we obtain the universal p-adic infinitesimal $\mathcal{O}_K$-thickening of $\mathcal{O}_{\mathbb{C}_p}$, usually denoted $\mathbf{A}_{\mathrm{inf}}$. Similarly, if we complete $\overline{K}$, we obtain $\mathbf{B}_{\mathrm{dR}}^+$, which is the universal p-adic infinitesimal $\mathcal{O}_K$-thickening of $\mathbb{C}_p$. We can obtain $\mathbf{B}_{\mathrm{dR}}^+$ from $\mathbf{A}_{\mathrm{inf}}$ by inverting p, and completing again.[12]

Alternatively, to get at the ring $\mathbf{A}_{\mathrm{inf}}$, we can directly consider a Grothendieck topology which allows infinitesimal thickenings. This will mean that global sections of sheaves on the corresponding site will need to extend compatibly on infinitesimal thickenings, a property which can be used to obtain rings which are universal infinitesimal thickenings. To do this, we define the (Zariski) **infinitesimal site**. Given a scheme X over S, this is the site $\mathrm{Inf}(X/S)$ whose objects are pairs $(U, \iota\colon U \to T)$ with U a Zariski open subset of X, and ι a closed immersion over S defined by a locally nilpotent sheaf of ideals,[13] exhibiting T as an infinitesimal thickening of U. The morphisms are then, of course, the expected commutative diagrams over S:

$$\begin{tikzcd} T_1 \arrow[r] & T_2 \\ U_1 \arrow[u, hook, "\iota_1"] \arrow[r] \arrow[dr, phantom, "\subseteq" rotate=-45] & U_2 \arrow[u, hook, "\iota_2"'] \\ & X \end{tikzcd}$$

Finally, the Grothendieck topology on this category is generated by families of morphisms $(\iota_i\colon U_i \hookrightarrow T_i) \to (\iota\colon U \hookrightarrow T)$ such that the morphisms $T_i \to T$ form a jointly surjective family of open immersions. We can thus define the **infinitesimal topos** $(X/S)_{\mathrm{inf}}$ of X over S as the corresponding topos of sheaves on $\mathrm{Inf}(X/S)$. As usual, the subcategory of abelian group objects in $(X/S)_{\mathrm{inf}}$ is an abelian category with enough injectives, and we

[12] Warning: describing $\mathbf{A}_{\mathrm{inf}}$ and $\mathbf{B}_{\mathrm{dR}}^+$ as universal infinitesimal deformations does not tell the whole story. This perspective is a good source of intuition about the purpose of these rings; however, explicit constructions carry with them additional information. For instance, the third construction, below, provides a lucid account of the action of the absolute Galois group of K, which will be crucial in applications.
[13] More generally, instead of nilpotent ideals one could consider the case of nil-ideals.

can thus define cohomology in this topos in the customary manner. In particular, we will consider the sheaf $\mathbb{O}_{X/S}\colon(\iota\colon U \hookrightarrow T) \mapsto \Gamma(T, \mathbb{O}_T)$. The foundational result pertaining to the infinitesimal topos is that its intrinsic cohomology computes algebraic de Rham cohomology *in characteristic* 0, so that if S is a $\mathbb{Q}$-scheme and $f\colon X \to S$ is smooth, there is a natural isomorphism [Grothendieck. Crystals, Theorem 4.1]

$$\mathrm{H}^i_{\mathrm{inf}}(X/S, \mathbb{O}_{X/S}) \cong \mathbf{H}^i_{\mathrm{Zar}}(X, \Omega^\bullet_{X/S}) = \mathrm{H}^i_{\mathrm{dR}}(X/S).$$

This relies on a version of the Poincaré lemma involving formal power series. We will see a more general version in section 5.3.2, which will also work in positive characteristic.
The infinitesimal topos also provides us with a direct relationship with notions of connections, and an immediate definition of the Gauss–Manin connection. We will cover this in more detail in section 5.3.3.

Now, to produce the rings with the required universal properties, we need to do this in levels: instead of directly forming infinitesimal $\mathcal{O}_K$-thickenings of $\mathcal{O}_{\mathbb{C}_p}$, we form successive infinitesimal $\mathcal{O}_K/(\varpi^n)$-thickenings of $\mathcal{O}_{\mathbb{C}_p}/(p)$. The purpose of this is twofold: it affords us the leisure to only deal with finite-order thickenings, and avoids us the need to consider the p-adic topological information. Doing it levelwise like this means that the whole process can be done algebraically, and the full picture is then recovered in the limit.[14]
Given Z over $\mathcal{O}_K/(\varpi^n)$, write $\mathbb{O}^{\mathrm{inf}}_n(Z) := \mathrm{H}^0_{\mathrm{inf}}\big(Z/(\mathcal{O}_K/(\varpi^n)), \mathbb{O}_{Z/(\mathcal{O}_K/(\varpi^n))}\big)$. The universal infinitesimal $\mathcal{O}_K/(\varpi^n)$-thickening of $\mathcal{O}_{\mathbb{C}_p}/(p^n)$ is then $\mathbb{O}^{\mathrm{inf}}_n(\mathcal{O}_{\bar{K}}/(p^n))$. Taking an inverse limit over n, we get $\mathbf{A}_{\mathrm{inf}} = \varprojlim_n \mathbb{O}^{\mathrm{inf}}_n(\mathcal{O}_{\bar{K}}/(p^n))$, the universal p-adic infinitesimal $\mathcal{O}_K$-thickening of $\mathcal{O}_{\mathbb{C}_p}$.

Finally, we could get our hands dirty and explicitly construct $\mathbf{B}^+_{\mathrm{dR}}$. The above approaches give us the idea of directly constructing an infinitesimal thickening. One possibility is to disregard all arithmetic information, and perform the violent act: $\mathcal{O}_{\mathbb{C}_p} \rightsquigarrow \mathcal{O}_{\mathbb{C}_p}[\![t]\!]$. This is not a good idea: we lose track of important information, such as Galois actions. Instead, we can perform a more arithmetically sensible operation: using Witt vectors in some form. The philosophy is that the Witt vector construction provides a universal infinitesimal thickening while keeping track of the Frobenius operator (this is a universal property enjoyed by Witt vectors, see for instance [Joyal. δ-rings, Théorème 4]). A plan of attack thus emerges: look at $\mathcal{O}_{\mathbb{C}_p}/(p)$, and render it perfect so as to apply the Witt vectors functor.

- We first start with $\mathcal{O}_{\mathbb{C}_p}/(p)$.
- We then take its perfection:

  $$\mathcal{O}_{\mathbb{C}_p^\flat} := \varprojlim_{x\mapsto x^p} \mathcal{O}_{\mathbb{C}_p}/(p) = \left\{(x^{(0)}, x^{(1)}, x^{(2)}, \dots) \in (\mathcal{O}_{\mathbb{C}_p}/(p))^{\mathbb{N}} : (x^{(i+1)})^p = x^{(i)}\right\}.$$

  In Colmez's notation, this ring is $\widetilde{\mathbf{E}}^+$.
- Next, we form the ring of (p-typical) Witt vectors $\mathcal{W}\left(\mathcal{O}_{\mathbb{C}_p^\flat}\right)$. This is $\mathbf{A}_{\mathrm{inf}}$, or in Colmez's notation, $\widetilde{\mathbf{A}}^+$.
- Finally, to get $\mathbf{B}^+_{\mathrm{dR}}$, we need to invert p and perform a certain (topological) completion (this completion can be described directly without using the open neighbourhoods $U_{n,k}$, as we will see shortly).

Let us now take the time to be more explicit.
For a start, the ring structure on the perfection of $\mathcal{O}_{\mathbb{C}_p}/(p)$ is defined as follows: given two sequences $x = (x^{(i)})_i$ and $y = (y^{(i)})_i$, we define $(x+y)^{(i)}$ as $\lim_{j\to+\infty}(x^{(i+j)} + y^{(i+j)})^{p^j}$, whereas for the product we can simply take a componentwise product. In addition, the reduction mod p morphism $\mathcal{O}_{\mathbb{C}_p} \to \mathcal{O}_{\mathbb{C}_p}/(p)$ yields an isomorphism $\varprojlim_{x\mapsto x^p} \mathcal{O}_{\mathbb{C}_p} \to \varprojlim_{x\mapsto x^p} \mathcal{O}_{\mathbb{C}_p}/(p)$.
This allows us then to define a map $\theta\colon \mathbf{A}_{\mathrm{inf}} \to \mathcal{O}_{\mathbb{C}_p}$ which sends $\sum_{i\geq 0} p^i [x_i]$ to $\sum_{i\geq 0} p^i x_i^{(0)}$, which is in fact a ring homomorphism. In terms of the first description of $\mathbf{A}_{\mathrm{inf}}$, θ simply forgets the data of the open sets $U_{n,k} = p^n \mathcal{O}^{(k)}_{\bar{K}}$

[14] We could bypass these problems by using different sites. For instance, we could replace the above small Zariski infinitesimal site by the big fppf infinitesimal site, which would solve the first problem (but not the second).

for $k > 0$, and so acts as a kind of decompletion. More rigourously, it sets equal to 0 all elements x such that $(p^{-i}x)_i$ remains bounded (in $\mathbf{A}_{\mathrm{inf}}[1/p]$).
With this explicit description of $\mathbf{A}_{\mathrm{inf}}$, we can precise the actions of the Galois group G_K and of the Frobenius endomorphism φ. Writing $x \in \mathbf{A}_{\mathrm{inf}}$ as $\sum_{i \geqslant 0} p^i [x_i]$, we have:

$$g \cdot \left(\sum_{i \geqslant 0} p^i [x_i] \right) = \sum_{i \geqslant 0} p^i [g \cdot x_i],$$

$$\varphi \left(\sum_{i \geqslant 0} p^i [x_i] \right) = \sum_{i \geqslant 0} p^i [\varphi(x_i)].$$

Moreover, we notice that $(\mathbf{A}_{\mathrm{inf}})^{\mathrm{G}_K} = \mathcal{W}\left(\left(\mathcal{O}_{\mathbb{C}_p^\flat}\right)^{\mathrm{G}_K}\right) = \mathcal{W}(\mathcal{O}_K/\mathfrak{m}_K) = W = \mathcal{O}_{K_0}$, because $\left(\mathbb{C}_p\right)^{\mathrm{G}_K} = K$ by the Ax–Sen–Tate theorem,[15] and also $(\mathbf{A}_{\mathrm{inf}})^{\varphi^r} = \mathcal{W}\left(\left(\mathcal{O}_{\mathbb{C}_p^\flat}\right)^{\varphi^r}\right) = \mathcal{W}(\mathbb{F}_{p^r}) = \mathbb{Z}_{p^r}$.

From $\mathbf{A}_{\mathrm{inf}}$, we then define $\mathbf{B}_{\mathrm{dR}}^+$ as the $\ker(\theta)$-adic completion of $\mathbf{A}_{\mathrm{inf}}[1/p]$:

$$\mathbf{B}_{\mathrm{dR}}^+ = \varprojlim_n \mathbf{A}_{\mathrm{inf}}[1/p]/\ker(\theta)^n.$$

The kernel of θ is principal, generated by any element $x \in \ker(\theta) \subset \mathbf{A}_{\mathrm{inf}}$ with $v_p(x_0^{(0)}) = 1$, where we recall that by definition of $\mathbf{A}_{\mathrm{inf}}$, we can write $x = (x_0, x_1, x_2, \ldots) \in \mathcal{W}\left(\mathcal{O}_{\mathbb{C}_p^\flat}\right)$, in which $x_i = (x_i^{(0)}, x_i^{(1)}, x_i^{(2)}, \ldots) \in \mathcal{O}_{\mathbb{C}_p^\flat}$, so that $x_0^{(0)}$ is an element of $\mathcal{O}_{\mathbb{C}_p}$ (and we can certainly take its p-adic valuation). To see why such an x generates $\ker(\theta) \subset \mathbf{A}_{\mathrm{inf}}$, it suffices to note that the inclusion $x\mathbf{A}_{\mathrm{inf}} \hookrightarrow \ker(\theta\colon \mathbf{A}_{\mathrm{inf}} \to \mathcal{O}_{\mathbb{C}_p})$ reduces mod p to a surjection $x_0\mathcal{O}_{\mathbb{C}_p^\flat} \to \ker\left(\theta\colon \mathcal{O}_{\mathbb{C}_p^\flat} \to \mathcal{O}_{\mathbb{C}_p}/(p)\right)$: this is because an element $y = (y^{(0)}, y^{(1)}, y^{(2)}, \ldots) \in \mathcal{O}_{\mathbb{C}_p^\flat}$ is in $\ker(\theta)$ if and only if $v_p(y^{(0)}) \geqslant 1$, and any such y is a multiple of x_0 as $v_p(x_0^{(0)}) = 1$. The result then follows by noting that both $\mathbf{A}_{\mathrm{inf}}$ and $\mathcal{O}_{\mathbb{C}_p}$ are p-adically complete, so that any morphism that reduces modulo p to a surjection is itself surjective by Nakayama's lemma.
From this, we deduce that $\mathbf{B}_{\mathrm{dR}}^+$ is a complete discrete valuation ring with respect to the $\ker(\theta)$-adic valuation; its residue field is $\mathbb{C}_p$. We already know that it contains K_0^{nr}, but in fact it contains all of $\overline{K}$ as we saw in the first description of $\mathbf{B}_{\mathrm{dR}}^+$. Alternatively we can just use Hensel's lemma for the map $\theta\colon \mathbf{B}_{\mathrm{dR}}^+ \to \mathbb{C}_p$. Note however that the inclusion $\overline{K} \to \mathbf{B}_{\mathrm{dR}}^+$ is *not* continuous for the p-adic topology of $\overline{K}$; indeed the crux of the first description of $\mathbf{B}_{\mathrm{dR}}^+$ was that we needed to add more open sets to the topology of $\overline{K}$ to get the subspace topology induced from the inclusion $\overline{K} \hookrightarrow \mathbf{B}_{\mathrm{dR}}^+$ (which were written $U_{n,k} = p^n \mathcal{O}_{\overline{K}}^{(k)}$).
A convenient generator of $\ker(\theta)$ one can take is $\xi = \widetilde{p} - p$, where $\widetilde{p} = [p^\flat] = (p^\flat, 0, 0, \ldots) \in \mathbf{A}_{\mathrm{inf}}$, and $p^\flat \in \mathcal{O}_{\mathbb{C}_p}^\flat$ is a compatible system of p-power roots of p.
We can think of $\mathbf{A}_{\mathrm{inf}}$ as a pretend 2-dimensional local ring, with coordinates given by p and $\tilde{p}$. Diagrammatically, we draw the specialisation poset of prime ideals of $\mathbf{A}_{\mathrm{inf}}$, labelling each point by the closed subscheme of $\operatorname{Spec}(\mathbf{A}_{\mathrm{inf}})$ it defines, and using arrows to denote inclusions of the corresponding subschemes (i.e. generisation maps).

[15] The Ax–Sen–Tate theorem states that, for any closed subgroup H of $\mathrm{G}_K \circlearrowright \overline{K}$ (also acting on $\mathbb{C}_p$ by continuity), $\left(\mathbb{C}_p\right)^H$ is the p-adic completion of $\overline{K}^H$.

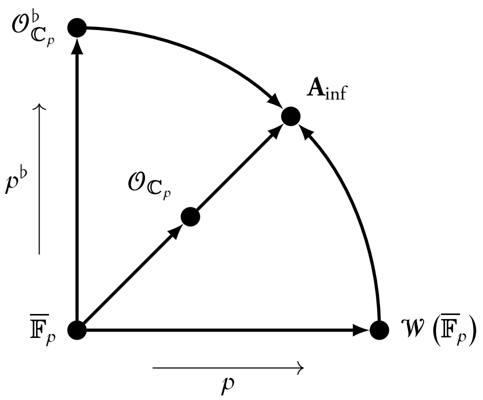


Figure I.1: Spec $(\mathbf{A}_{\mathrm{inf}})$.

Another particularly interesting element in $\ker(\theta)$ comes from our previous considerations for the monodromy of the logarithm (section 2). We were interested in the sequence $(p^n \log(\varepsilon_n))$ of elements of $\mathbb{C}_p$, where ε_n formed a compatible system of primitive p^n-th roots of unity. In this new context, define $\tilde{\varepsilon}_n = [(\varepsilon_n, \varepsilon_{n+1}, \ldots)]$, the Teichmüller lift of the sequence $(\varepsilon_n, \varepsilon_{n+1}, \ldots)$ in $\mathcal{O}_{\mathbb{C}_p^\flat}$. These $\tilde{\varepsilon}_n$ are elements of $\mathbf{A}_{\mathrm{inf}}$, with $\theta(\tilde{\varepsilon}_n) = \varepsilon_n$. We now have a better shot at defining a non-zero limit of the sequence $(p^n \log(\tilde{\varepsilon}_n))_n$ in $\mathbf{B}^+_{\mathrm{dR}}$. It will suffice to define $\log(\tilde{\varepsilon}_0)$, as then we will have $\log(\tilde{\varepsilon}_n) = p^{-n} \log(\tilde{\varepsilon}_0)$. For simplicity, write $\tilde{\varepsilon}$ for $\tilde{\varepsilon}_0$. Contemplating the formal expression

$$\log(\tilde{\varepsilon}) = -\sum_{n \geqslant 1} \frac{(1-\tilde{\varepsilon})^n}{n},$$

the observation that $1 - \tilde{\varepsilon}$ is in the kernel of θ allows us to ascertain the convergence of this sum in $\mathbf{B}^+_{\mathrm{dR}}$. In fact, one can see this as a justification for the necessity of completing $\ker(\theta)$-adically after inverting p in $\mathbf{A}_{\mathrm{inf}}$.
This element $\log(\tilde{\varepsilon})$ is often denoted t, and is Fontaine's *p-adic analogue* of $2\pi\mathrm{i}$. It is the p-adic monodromy of the logarithm, i.e. a period for the p-adic cyclotomic character,[16] so that $g \cdot \mathrm{t} = \chi_p(g)\mathrm{t}$, with χ_p the p-adic cyclotomic character. We can write the p-adic Euler identity: $\exp(\mathrm{t}) = \tilde{\varepsilon}$. Interestingly, $\tilde{\varepsilon} \neq 1$ in $\mathbf{B}^+_{\mathrm{dR}}$, although its image in $\mathbb{C}_p$ is $\varepsilon_0 = 1$.
Finally, we define $\mathbf{B}_{\mathrm{dR}} := \mathbf{B}^+_{\mathrm{dR}}[1/\mathrm{t}]$. This is Fontaine's **ring of de Rham periods**, which is the hoped for receptacle for the p-adic integral of de Rham cohomology classes on varieties over K. As t is a uniformiser of $\mathbf{B}^+_{\mathrm{dR}}$, this just means that $\mathbf{B}_{\mathrm{dR}}$ is the fraction field of $\mathbf{B}^+_{\mathrm{dR}}$. Using $\mathrm{t} \in \mathbf{B}^+_{\mathrm{dR}}$, we have an embedding of $\mathbb{Q}_p(1)$[17] into $\mathbf{B}^+_{\mathrm{dR}}$ (as we saw that G_K acts on t via χ_p). Moreover, we can define a filtration $F^\bullet$ on $\mathbf{B}_{\mathrm{dR}}$ by declaring $F^i\mathbf{B}_{\mathrm{dR}} = \mathrm{t}^i\mathbf{B}^+_{\mathrm{dR}}$, which is then stable under G_K. This is the **Hodge filtration** on $\mathbf{B}_{\mathrm{dR}}$.
Unfortunately, even though we were careful to use the Witt vector construction in order to bring along the Frobenius endomorphism, it nevertheless got lost on the way to $\mathbf{B}_{\mathrm{dR}}$. This is because the natural Frobenius of $\mathbf{A}_{\mathrm{inf}}$ does not preserve $\ker(\theta)$: $\varphi(\xi) = (\xi + p)^p - p \notin \ker(\theta)$. We will strive to remedy this in section 5.4.

We can also extend our computation of $(\mathbf{A}_{\mathrm{inf}})^{\mathrm{G}_K} = K$ to the computation of $(\mathbf{B}_{\mathrm{dR}})^{\mathrm{G}_K}$. No new invariants were added in the process of completing $\ker(\theta)$-adically:

Proposition 3.1 — The Galois invariants of $\mathbf{B}_{\mathrm{dR}}$ are given by $(\mathbf{B}_{\mathrm{dR}})^{\mathrm{G}_K} = (\mathbf{B}^+_{\mathrm{dR}})^{\mathrm{G}_K} = K$. ◇

Proof: We take Galois invariants in the following exact sequence

$$0 \longrightarrow \mathrm{t}^{i+1}\mathbf{B}^+_{\mathrm{dR}} \longrightarrow \mathrm{t}^i\mathbf{B}^+_{\mathrm{dR}} \xrightarrow{\theta} \mathbb{C}_p(i) \longrightarrow 0.$$

[16] Note that, just as $2\pi\mathrm{i}$ depends on a choice of orientation on $\mathbb{C}$, this element t depends on a choice of p-adic orientation, as it depends on ε.
[17] Here $\mathbb{Q}_p(1)$ denotes a Tate twist of $\mathbb{Q}_p$. We recall that $\mathbb{Q}_p(i)$ is a 1-dimensional $\mathbb{Q}_p$-vector space on which G_K acts via χ_p^i.

We thus obtain

$$\begin{tikzcd} 0 \arrow[r] & (t^{i+1}\mathbf{B}_{\mathrm{dR}}^+)^{G_K} \arrow[r] & (t^i\mathbf{B}_{\mathrm{dR}}^+)^{G_K} \arrow[r] & (\mathbb{C}_p(i))^{G_K} \arrow[dll, rounded corners] \\ & H^1(G_K, t^{i+1}\mathbf{B}_{\mathrm{dR}}^+) \arrow[r] & H^1(G_K, t^i\mathbf{B}_{\mathrm{dR}}^+) \arrow[r] & H^1(G_K, \mathbb{C}_p(i)). \end{tikzcd}$$

The Ax–Sen–Tate theorem then allows us to describe the rightmost column; it implies that $H^n(G_K, \mathbb{C}_p(i)) = 0$ for $i \neq 0$ and all n. This shows that for $i \neq 0$, $(t^{i+1}\mathbf{B}_{\mathrm{dR}}^+)^{G_K} = (t^i\mathbf{B}_{\mathrm{dR}}^+)^{G_K}$ and $H^1(G_K, t^{i+1}\mathbf{B}_{\mathrm{dR}}^+) = H^1(G_K, t^i\mathbf{B}_{\mathrm{dR}}^+)$. We deduce from the former that $(t^{-i}\mathbf{B}_{\mathrm{dR}}^+)^{G_K} = (\mathbf{B}_{\mathrm{dR}}^+)^{G_K}$ for any $i \geqslant 0$. As $\mathbf{B}_{\mathrm{dR}} = \bigcup_{i\geqslant 0} t^{-i}\mathbf{B}_{\mathrm{dR}}^+$, we obtain the equality $(\mathbf{B}_{\mathrm{dR}})^{G_K} = (\mathbf{B}_{\mathrm{dR}}^+)^{G_K}$.
Considering the case $i = 0$, by the Ax–Sen–Tate theorem we are left with the exact sequence

$$0 \to (t\mathbf{B}_{\mathrm{dR}}^+)^{G_K} \to (\mathbf{B}_{\mathrm{dR}}^+)^{G_K} \to K \to H^1(G_K, t\mathbf{B}_{\mathrm{dR}}^+),$$

so that the desired result $(\mathbf{B}_{\mathrm{dR}}^+)^{G_K} = K$ is seen to follow from the two assertions $(t\mathbf{B}_{\mathrm{dR}}^+)^{G_K} = 0$ and $H^1(G_K, t\mathbf{B}_{\mathrm{dR}}^+) = 0$. These respectively follow from $(t^{i+1}\mathbf{B}_{\mathrm{dR}}^+)^{G_K} = (t^i\mathbf{B}_{\mathrm{dR}}^+)^{G_K}$ and $H^1(G_K, t^{i+1}\mathbf{B}_{\mathrm{dR}}^+) = H^1(G_K, t^i\mathbf{B}_{\mathrm{dR}}^+)$, for $i > 0$. □

4 The de Rham comparison theorem

With $\mathbf{B}_{\mathrm{dR}}$ in hand, we hope to be able to perform p-adic integration. As we're working purely with cohomology (and not with homology), we expect something that mirrors the isomorphism

$$H^n_{\mathrm{sing}}(X(\mathbb{C}), K) \otimes_K \mathbb{C} \cong H^n_{\mathrm{dR}}(X/K) \otimes_K \mathbb{C}.$$

of section 1.4. This is the content of the following theorem, conjectured by Fontaine [Fontaine. $\mathbf{B}_{\mathrm{cris}}$, Conjecture A.6]:

Theorem 4.1 (Faltings) — Let X be a smooth proper variety over K. There exists a natural, functorial isomorphism

$$C_{\mathrm{dR}}\colon H^i_{\text{ét}}\left(X_{\overline{K}}, \mathbb{Q}_p\right) \otimes_{\mathbb{Q}_p} \mathbf{B}_{\mathrm{dR}} \xrightarrow{\sim} H^i_{\mathrm{dR}}(X/K) \otimes_K \mathbf{B}_{\mathrm{dR}}.$$ [18]

Furthermore, this isomorphism is G_K-equivariant, respects the filtrations defined on both sides, and is compatible with Poincaré duality, the Künneth formula, Tate twists, the formation of Chern classes of vector bundles, and cycle class maps.[19] ◇

Proof: The original proof is due to Faltings, and doesn't explicitly involve any notion of p-adic integration. In the case of H^1, it is however possible to perform integration on abelian varieties to prove this result [Colmez. $\int$, Théorème II.3.2]. Yet another method of proof is to reduce to the semistable comparison theorem, which we will see in section 10. □

Now, as this comparison isomorphism preserves the Galois action and the filtrations, and as $(\mathbf{B}_{\mathrm{dR}})^{G_K} = K$ by Proposition 3.1, this means we can recover $H^i_{\mathrm{dR}}(X/K)$ as a filtered K-vector space from $H^i_{\text{ét}}\left(X_{\overline{K}}, \mathbb{Q}_p\right)$ by the following procedure:

$$H^i_{\mathrm{dR}}(X/K) \cong \left(H^i_{\text{ét}}\left(X_{\overline{K}}, \mathbb{Q}_p\right) \otimes_{\mathbb{Q}_p} \mathbf{B}_{\mathrm{dR}}\right)^{G_K}.$$

Unfortunately we cannot expect to go back the other way, as the Hodge filtration on the de Rham cohomology is certainly not enough to recover the Galois action on the étale cohomology.

[18] This also extends to an isomorphism of E_∞-algebras $C_{\mathrm{dR}}\colon R\Gamma_{\text{ét}}(X_{\overline{K}}, \mathbb{Q}_p) \otimes_{\mathbb{Q}_p} \mathbf{B}_{\mathrm{dR}} \xrightarrow{\sim} \mathbf{R}\Gamma_{\mathrm{Zar}}(X, \Omega^\bullet_{X/K}) \otimes_K \mathbf{B}_{\mathrm{dR}}$, compatibly with all the extra structures.
[19] One could say it is an isomorphism of Bloch–Ogus cohomology theories.

5 The ring $\mathrm{B}_{\mathrm{cris}}^{+}$

With C_{dR}, we were considering a general smooth projective variety X over K. However, one special case we might want to consider is when X has a smooth model $\mathcal{X}$ over $\mathcal{O}_K$. Then, on top of the cohomology of the generic fibre X, we can also try to bring into play the cohomology of the special fibre $\mathcal{X}_\kappa = \mathcal{X} \times_{\mathcal{O}_K} \kappa$. We might then hope to transfer the problem of calculating periods of varieties over K to the hopefully more manageable problem of computing periods of varieties over κ. To do this we would like to work with cohomology theories over κ; however, as κ has positive characteristic, this poses some problems...

5.1 Weil cohomology theories

We're interested in the cohomological invariants of the special fibre $Y = \mathcal{X}_\kappa$. To that end, suppose that more generally we are dealing with a smooth projective variety Y over an arbitrary field κ of characteristic p. What do we have?

First off, we have the étale cohomology of Y, $\mathrm{H}^i_{\text{ét}}(Y, \mathcal{F})$. This is most relevant in the situation when $\mathcal{F}$ is a constructible sheaf. The obvious choice, in order to obtain a cohomology with coefficients resembling coefficients in $\mathbb{Z}$, is to look at coefficients of the form $\mathbb{Z}/N\mathbb{Z}$. Looking one prime at a time, we end up defining the ℓ-adic étale cohomology $\mathrm{H}^i_{\text{ét}}(Y, \mathbb{Z}_\ell) := \varprojlim_N \mathrm{H}^i_{\text{ét}}(Y, \mathbb{Z}/\ell^N\mathbb{Z})$.[20] This has adequate properties when $\ell \neq p$: the dimensions correspond to what we expect in light of the behaviour of varieties over $\mathbb{C}$ (as indicated by the singular cohomology of their set of complex points), we have results of finite dimensionality (recall: for smooth projective Y), etc. However, the situation is less idyllic when $\ell = p$.
Indeed, recall that the ℓ-adic étale cohomology for curves is usually computed using the following basic ingredients:

- The computation of the étale cohomology of $\mathrm{Spec}(K)$, for K a field, using Galois cohomology through Grothendieck's Galois theory (which identifies Čech cohomology for the étale topos with the corresponding Galois cohomology).
- A comparison of the étale and Zariski toposes. The morphism of toposes, from the étale topos to the Zariski topos, induces an isomorphism in cohomology with coefficients in quasicoherent sheaves. That is, we have canonical isomorphisms $\mathrm{H}^i_{\text{ét}}(Y, \mathcal{M}) \to \mathrm{H}^i_{\mathrm{Zar}}(Y, \mathcal{M})$ for any quasicoherent sheaf of $\mathcal{O}_Y$-modules $\mathcal{M}$. This is a result of the Leray spectral sequence, and the vanishing of higher direct images along this morphism for quasicoherent sheaves. See, for instance, [SGA 4_{II}, Exposé VII, Proposition 4.3].
- The Kummer short exact sequence of sheaves on Y:
$$0 \to \mu_{\ell^N} \longrightarrow \mathbb{G}_{\mathrm{m}} \xrightarrow{x \mapsto x^{\ell^N}} \mathbb{G}_{\mathrm{m}} \to 0.$$
When $p \neq \ell$, this is a sequence of étale sheaves (as then the polynomials $T^{\ell^N} - a$ are all separable).
- The computation of $\mathrm{H}^1_{\text{ét}}(Y, \mathbb{G}_m) = \mathrm{Pic}(Y)$, the Picard group of Y.

When $\ell = p$, the sequence
$$0 \to \mu_{\ell^N} \longrightarrow \mathbb{G}_{\mathrm{m}} \xrightarrow{x \mapsto x^{\ell^N}} \mathbb{G}_{\mathrm{m}} \to 0$$
is no longer étale. In its stead, we have the Artin–Schreier sequence
$$0 \to \mathbb{Z}/p\mathbb{Z} \longrightarrow \mathbb{G}_{\mathrm{a}} \xrightarrow{x \mapsto x^p - x} \mathbb{G}_{\mathrm{a}} \to 0.$$
Now, we can compute the cohomology of $\mathbb{G}_{\mathrm{a}}$ in the étale topos as the cohomology in the Zariski topos, as $\mathcal{O}_Y$ is coherent. Therefore, as $\mathrm{H}^i_{\mathrm{Zar}}(Y, \mathcal{O}_Y)$ vanishes for $i \notin [0, d]$ (where $d = \dim(Y)$), the same applies to $\mathrm{H}^i_{\text{ét}}(Y, \mathcal{O}_Y)$, and therefore also for $\mathrm{H}^i_{\text{ét}}(Y, \mathbb{Z}/p\mathbb{Z})$ and thus too for the p-adic étale cohomology of Y, $\mathrm{H}^i_{\text{ét}}(Y, \mathbb{Z}_p)$.

[20] Alternatively, we could do this in one step by considering cohomology in the pro-étale topos.

Clearly then, in characteristic p, p-adic étale cohomology is too stunted for our interests.

How about, then, using the algebraic de Rham cohomology of Y, as we saw in section 1.2? Now, in characteristic p, it is no longer necessarily true that the Hodge-to-de Rham spectral sequence degenerates at E_1, and we cannot expect to have opposed filtrations either, as we saw. Moreover, $H^i_{dR}(Y/\kappa)$ is a κ-vector space, whereas it would have been preferable to have an object in characteristic 0 in order to have a Lefschetz fixed point formula like one has in ℓ-adic cohomology. Yet, these are not particularly serious defects; they point towards interesting geometric behaviour of varieties in characteristic p.

An illuminating consideration is to go back to the calculation of the algebraic de Rham cohomology using the infinitesimal site that we saw in section 3, with the second description of $\mathbf{B}^+_{dR}$. This said that for X/K a smooth scheme over a field of characteristic 0, the formal Poincaré lemma gave us a canonical isomorphism between $H^i_{inf}(X/K, \mathcal{O}_{X/K})$ and $H^i_{dR}(X/K)$. This runs into problems in characteristic p: we no longer dispose of a Poincaré lemma *at all*. Nothing can save the situation: not an algebraic Poincaré lemma, not an analytic Poincaré lemma, not a formal Poincaré lemma, ...
The reason is simple: $d(x^p) = px^{p-1}dx = 0$ in characteristic p. This means we are not going to be able to find a primitive for x^{p-1}, even formally. In particular, this implies that $H^1_{dR}(\mathbb{A}^1/\kappa)$ is infinite dimensional. Again, not ideal.

5.2 Divided powers

How do we fix this? To obtain a formal Poincaré lemma, we require the ability to manipulate Taylor series. Formally, we want a formula resembling:

$$f(x+\varepsilon) = \sum_{n=0}^{+\infty} f^{(n)}(x) \cdot \frac{\varepsilon^n}{n!}.$$

We can investigate the nature of the coefficients $\frac{\varepsilon^n}{n!}$ to come up with a definition of the kind of structure necessary on a ring to perform Taylor expansions. Say we have a ring R with a collection of infinitesimal elements I, and want coefficients $\gamma_n(\varepsilon)$ for $\varepsilon \in I$ with a notion of Taylor expansion:

$$f(x+\varepsilon) = \sum_{n=0}^{+\infty} f^{(n)}(x) \cdot \gamma_n(\varepsilon).$$

We can deduce some necessary properties of such a structure already:

- I should be an ideal. This is immediate from the notion of an infinitesimal. Furthermore, $\gamma_n(\varepsilon)$ should also be infinitesimal for $n \geqslant 1$, as all terms in the Taylor expansion of $f(x+\varepsilon)$, except the constant term, should vanish when we take $\varepsilon = 0$.
- Taking f to be a constant function, we must have that $\gamma_0(\varepsilon) = 1$ for all $\varepsilon \in I$.
- Taking f to be the identity function, we get that $\gamma_1(\varepsilon) = \varepsilon$ for all $\varepsilon \in I$.
- The Taylor expansion of $f(x + r\varepsilon)$ shows that we must have $\gamma_n(r\varepsilon) = r^n\gamma_n(\varepsilon)$.
- Considering the Taylor expansion of $f(x+\varepsilon+\eta)$ for $\varepsilon, \eta \in I$, we expect $\gamma_n(\varepsilon+\eta) = \sum_{i+j=n} \gamma_i(\varepsilon)\gamma_j(\eta)$.
- From the product rule, we see that we should have $\gamma_m(\varepsilon)\gamma_n(\varepsilon) = \frac{(m+n)!}{m!n!}\gamma_{m+n}(\varepsilon)$.[21]
- From the chain rule, we obtain the formula $\gamma_m(\gamma_n(\varepsilon)) = \frac{m!n!}{m!(n!)^m}\gamma_{mn}(\varepsilon)$.[22]

[21] Note that $\frac{(m+n)!}{m!n!}$ is an integer: it is the binomial coefficient $\binom{n+m}{n,m}$.
[22] Of course, $\frac{m!n!}{m!(n!)^m}$ is also an integer: it counts arrangements of mn labelled balls into m groups of n.

This motivates the definition of a ring with a **divided power structure** (or **PD-structure**, from the French "puissances divisées"). This consists of a ring R, an ideal I of R, and a collection of functions $\gamma_0 \colon I \to R$, $\gamma_n \colon I \to I$ satisfying the above conditions.
Note that by iterating the product rule formula, we obtain that $n!\gamma_n(\varepsilon) = \varepsilon^n$. When $n!$ is invertible in R, this means we recover the previous formula $\gamma_n(\varepsilon) = \frac{\varepsilon^n}{n!}$. If $n!$ is only known to not be a zero-divisor, then any PD-structure, if it exists, must satisfy $\gamma_n(\varepsilon) = \frac{\varepsilon^n}{n!}$.

We come back to our bread and butter: take $R = \mathcal{O}_K$ to be the ring of integers of a finite extension of $\mathbb{Q}_p$. From the previous remarks, a PD-structure on $\mathcal{O}_K$ (with respect to its maximal ideal (ϖ)), if it exists, must be given by $\gamma_n(\varpi) = \frac{\varpi^n}{n!}$. We need to know when this defines a valid PD-structure. The only property we need to check is that indeed $\gamma_n(\varpi)$ is an element of the maximal ideal of $\mathcal{O}_K$. It suffices to compute its valuation (which is given by Legendre's formula), to check it is at least 1. Writing the base p expansion of n as $n = \sum_i a_i p^i$, we have:

$$\mathrm{v}_p(n!) = \frac{n - \sum_i a_i}{p-1}.$$

Now, by definition of the ramification index e, we have that $\varpi^e = p$, so that:

$$\mathrm{v}_\varpi(\gamma_n(\varpi)) = \mathrm{v}_\varpi(\varpi^n) - \mathrm{v}_\varpi(n!) = n - e\mathrm{v}_p(n!) = n\frac{p-1-e}{p-1} + e\frac{\sum_i a_i}{p-1}.$$

This is a decreasing function of e, which for $e = p-1$ is equal to $\sum_i a_i \geqslant 1$ (for $n > 0$). For $e = p$ and $n = p$ the expression evaluates to $0 \ngeqslant 1$. Therefore $\mathcal{O}_K$ admits a PD-structure with respect to its maximal ideal if and only if $e \leqslant p-1$. This structure is unique, by torsion-freeness of R.

Note that we can extend the notions defined in this section to schemes, leading to the notion of **PD-schemes**.

5.3 The crystalline topos

We hope now to adjust the definition of the infinitesimal topos of section 3 to take into account the subtleties of the situation in positive characteristic.

The notion replacing infinitesimal thickenings is that of PD-thickenings. To define these, we need to consider morphisms of PD-schemes. At the affine level, there's the obvious notion of a homomorphism of PD-rings, i.e. a morphism $f \colon (A, I, \gamma) \to (B, J, \delta)$ with $IB \subseteq J$ and $f \circ \gamma_i = \delta_i \circ f$. This notion then leads to the definition of morphisms of PD-schemes. However it is fruitful to relax the condition $IB \subseteq J$. When that condition no longer necessarily holds, we say f is *compatible with the PD-structures* if there is a divided power structure ε on the ideal $J + IB$ such that the corresponding homomorphisms $f' \colon (A, I, \gamma) \to (B, J + IB, \varepsilon)$ and $g \colon (B, J, \delta) \to (B, J + IB, \varepsilon)$ are morphisms of PD-rings, i.e. $f' \circ \gamma_i = \varepsilon_i \circ f'$, $g \circ \delta_i = \varepsilon_i \circ f'$. Note that such a PD-structure ε, if it exists, is unique [Berthelot. $\mathrm{H}_{\mathrm{cris}}$, Lemme I.2.2]. The definition then extends in the obvious manner to morphisms of schemes.

We can now envision how to adapt the notion of infinitesimal thickenings to the case of PD-ideals. A PD-ideal (I, γ) is said to be **PD-nilpotent** if $I^{[n]} = 0$ for some natural number n, where $I^{[n]}$ is the ideal generated by elements of the form $\gamma_{i_1}(x_1) \cdots \gamma_{i_k}(x_k)$ for $i_1 + \cdots + i_k \geqslant n$. It is said to be **quasi-PD-nilpotent** if $mI^{[n]} = 0$ for some natural numbers m, n. This is the correct notion if we want the theory to work both in characteristic 0 and in characteristic p [Berthelot. $\mathrm{H}_{\mathrm{cris}}$, Appendice]. In characteristic 0 this reduces both notions of PD-nilpotence and quasi-PD-nilpotence to that of the ideal being nilpotent. On the other hand, in characteristic p, or more generally if p is nilpotent, these imply that the ideal is nilpotent, respectively is a nil-ideal [Berthelot. $\mathrm{H}_{\mathrm{cris}}$, Proposition I.3.1.4]. This then adequately generalises the situation for the infinitesimal site, and we define a **PD-thickening** to be a closed immersion of schemes, defined by a quasi-coherent sheaf of locally quasi-PD-nilpotent ideals, which is compatible with the PD-structures.

Consider now a PD-scheme $S = (\underline{S}, \mathcal{I}, \gamma)$, and $f: Y \to \underline{S}$ a scheme over $\underline{S}$, equipped with a PD-structure compatible with γ (such a PD-structure is uniquely determined when it exists, as observed above). The (Zariski) **crystalline site** $\mathrm{Cris}(Y/S)$ of Y over S is then defined as follows. Its objects consist of Zariski open subsets $U \subseteq Y$ together with PD-thickenings (ι, δ), with $\iota: U \to T$ a closed immersion and δ the associated PD-structure. The compatibility is required with respect to the PD-structure on U inherited from Y. Morphisms are commutative diagrams of PD-thickenings:

$$\begin{tikzcd} T_1 \arrow[rr, "g"] & & T_2 \\ U_1 \arrow[u, hook, "\iota_1"] \arrow[rr] \arrow[dr, phantom, sloped, "\subseteq"] & & U_2 \arrow[u, hook, "\iota_2"'] \arrow[dl, phantom, sloped, "\supseteq"] \\ & Y & \end{tikzcd}$$

Note that we require here that g is a morphism of PD-schemes, not just that it be compatible with the PD-structures. Finally, the Grothendieck topology is generated by families of morphisms $(U_i, T_i, \delta_i) \to (U, T, \delta)$ such that the morphisms $T_i \to T$ are a jointly surjective family of open immersions. The associated **crystalline topos** of sheaves on this site is then written $(Y/S)_{\mathrm{cris}}$. As before, it is appropriate to consider the sheaf $\mathcal{O}_{Y/S}: (U, \iota: U \hookrightarrow T, \delta) \mapsto \Gamma(T, \mathcal{O}_T)$. Other important sheaves are $\mathbb{G}_{\mathrm{a},Y/S}: (U, \iota: U \hookrightarrow T, \delta) \mapsto \Gamma(U, \mathcal{O}_U)$, and $\mathcal{J}_{Y/S} = \ker(\mathcal{O}_{Y/S} \to \mathbb{G}_{\mathrm{a}})$. Moreover, $\mathcal{J}_{Y/S}$ is naturally endowed with a PD-structure, precisely because each $U \hookrightarrow T$ is a defined by a PD-ideal sheaf.

One significant gain in using the topos-theoretic language is that we have a functoriality result: the inverse image of a (PD-)thickening might not be a (PD-)thickening, but we can always consider the (non-necessarily representable) sheaf defined by inverse image of the sheaf representing the (PD-)thickening. In this manner, given a commutative square

$$\begin{tikzcd} Z \arrow[r, "h"] \arrow[d, "g"'] & Y \arrow[d, "f"] \\ (R, \mathcal{I}, \gamma) \arrow[r, "r"] & (S, \mathcal{J}, \delta), \end{tikzcd}$$

with r a morphism of PD-schemes, there is a corresponding morphism of (PD-ringed) toposes $h_{\mathrm{cris}}: (Z/R)_{\mathrm{cris}} \to (Y/S)_{\mathrm{cris}}$ [Berthelot. $\mathrm{H}_{\mathrm{cris}}$, Corollaire III.2.2.4].[23]

Associated to this topos, of course, is its intrinsic cohomology theory, **crystalline cohomology**. For instance, we can consider the crystalline cohomology of the structure sheaf $\mathrm{H}^i_{\mathrm{cris}}(Y/S, \mathcal{O}_{Y/S})$.
However, we are most interested in the following set-up: κ a perfect field of characteristic p, with $W = \mathcal{W}(\kappa)$ its ring of p-typical Witt vectors, with truncations $W_n = W/p^{n+1}W$. The ring of Witt vectors W is equipped with its canonical (and unique) PD-structure on its maximal ideal as seen in section 5.2, and W_n with its quotient PD-structure (which is not the only PD-structure on W_n). We can work either directly with W, or if we prefer to use PD-nilpotent ideals instead of only quasi-PD-nilpotent ideals, the rings W_n in which p is nilpotent; the latter avoids several technical complications such as having to work with formal schemes and topologically nilpotent elements. Given then a smooth proper scheme $\mathcal{X}/\mathcal{O}_K$, with generic fibre X/K and special fibre Y/κ, we define

$$\mathrm{H}^i_{\mathrm{cris}}(Y/W, \mathcal{O}_{Y/W}) = \varprojlim_n \mathrm{H}^i_{\mathrm{cris}}(Y/W_n, \mathcal{O}_{Y/W_n}),$$

the crystalline cohomology of Y/W. Note that Y is naturally a W-scheme by virtue of the quotient homomorphism $W \to \kappa$; this makes Y a scheme over W with only a special fibre, and empty generic fibre! Transport of structure for the absolute Frobenius endomorphism Fr_Y on $(Y/W_n)_{\mathrm{cris}}$ then naturally yields a Frobenius endomorphism φ acting on these crystalline cohomology groups, which is semilinear with respect to the Frobenius

[23] We can describe the backward portion of this morphism explicitly: let $h^*(U_1, T_1, \iota_1)$ be the sheaf that to $(U_2, T_2, \iota_2) \in \mathrm{Cris}(Z/R)$ associates the set of all PD-homomorphisms $(T_2, \iota_2) \to (T_1, \iota_1)$ commuting with $h: U_2 \to U_1$ (assuming that $h(U_2) \subseteq U_1$, otherwise just associate to (U_2, T_2, ι_2) the empty set) over $r: R \to S$. Once h^* defined, there is a unique way to define h_* to get a morphism of toposes [Berthelot. $\mathrm{H}_{\mathrm{cris}}$, Théorème III.2.2.3].

endomorphism σ that acts on the coefficients W.

5.3.1 Crystals

The difficulty with the crystalline topos – and the infinitesimal topos – is that the terminal object is not representable, whereas for Y_{Zar} or $Y_{\mathrm{\acute{e}t}}$ for instance, the terminal object $\mathbf{1}$ is represented by Y itself. The issue is that in $(Y/S)_{\mathrm{cris}}$, on top of taking smaller and smaller $U \subseteq Y$, we can also take increasing PD-thickenings $U \hookrightarrow T$ of U.[24]

To get a handle on the crystalline topos, therefore, it is useful to use the forgetful functor

$$\mathrm{u}_{Y/S} \colon \mathrm{Cris}(Y/S) \to \mathrm{Zar}(Y/S), \qquad \mathrm{u} \colon (U, T, \delta) \mapsto U,$$

which we'll often also write simply u.

This morphism of sites then induces a geometric morphism of the corresponding toposes

$$\begin{tikzcd} (Y/S)_{\mathrm{cris}} \arrow[r, bend left, "\mathrm{u}_*"] \arrow[r, phantom, "\top"] & (Y/S)_{\mathrm{Zar}} \arrow[l, bend left, "\mathrm{u}^*"] \end{tikzcd}.$$

The benefit is that the pushforward morphism from $(Y/S)_{\mathrm{cris}}$ to the "topos ponctuel" **Set** factors through u_*, meaning that we can compute crystalline cohomology as hypercohomology in the Zariski topos

$$\mathrm{H}^i_{\mathrm{cris}}\left(Y/S, \mathcal{O}_{Y/S}\right) = \mathbf{H}^i_{\mathrm{Zar}}\left(Y, \mathrm{Ru}_*\mathcal{O}_{Y/S}\right).$$

We can also describe all crystalline sheaves using Zariski sheaves. Given a sheaf $\mathcal{F} \in (Y/S)_{\mathrm{cris}}$ and a morphism $g \colon (U_1, T_1, \delta_1) \to (U_2, T_2, \delta_2)$ in $\mathrm{Cris}(Y/S)$, there is a transition morphism $g^* \colon g^{-1}\mathcal{F}|_{U_2,T_2,\delta_2} \to \mathcal{F}|_{U_1,T_1,\delta_1}$. These transition morphisms form a compatible system, in that

$$(h \circ g)^* = g^* \circ g^{-1}(h^*).$$

This means that we can describe $\mathcal{F}$ as a compatible system of Zariski sheaves. In general, given such a system of Zariski sheaves, with the condition that g^* is an isomorphism whenever g is an open immersion, there is a unique sheaf on the crystalline site that is described by this system of Zariski sheaves [Berthelot. $\mathrm{H}_{\mathrm{cris}}$, III.1.1.4].

One particular important category of sheaves on the crystalline site is that of **crystals**, which parallels the notion of quasi-coherent modules for the Zariski topos. The simplest crystal is the structure sheaf $\mathcal{O}_{Y/S}$ itself. More generally, one has crystals of $\mathcal{O}_{Y/S}$-modules: these are described by a Zariski system of $\mathcal{O}_{Y/S}$-modules, i.e. consist of, for each $(U, T, \delta) \in \mathrm{Cris}(Y/S)$, a sheaf $\mathcal{M}_{(U,T,\delta)}$ of $\mathcal{O}_T$-modules on the Zariski site of T, such that the transition morphisms

$$g^\sharp \colon g^*\mathcal{M}_{(U_2,T_2,\delta_2)} \to \mathcal{M}_{(U_1,T_1,\delta_1)}$$

are isomorphisms, where $g^*\mathcal{M}_{(U_2,T_2,\delta_2)} = g^{-1}\mathcal{M}_{(U_2,T_2,\delta_2)} \otimes_{g^{-1}\mathcal{O}_{T_1}} \mathcal{O}_{T_2}$ is the $\mathcal{O}_{Y/S}$-module pullback. To capture related notions (such as crystals in $\mathcal{O}_{Y/S}$-algebras, crystals in schemes, and so on), we can replace the fibred category of $\mathcal{O}_{Y/S}$-modules with an arbitrary fibred category over the crystalline site [Berthelot. $\mathrm{H}_{\mathrm{cris}}$, Définition IV.1.1.1]. Just like one can use different Grothendieck topologies to compute the cohomology of quasi-coherent modules, crystals have the same cohomology whether one uses the Zariski, étale or fppf topologies to build the crystalline topos out of.

The fundamental property of crystals is that they can grow along infinitesimal thickenings.[25] For instance, we can consider the thickenings of $U \subseteq Y$ given by infinitesimal neighbourhoods of the diagonal of U in $U \times_S U$. This allows us to relate crystals with connections! The basic idea is that, instead of asking for parallel transport along paths (which are not an algebraic notion), we should be able to compatibly parallel transport to infinitesimally

[24] However, when Y embeds as a closed subscheme of a smooth S-scheme Z, the terminal object can be covered by a representable sheaf: the structure sheaf of the PD-envelope of Y in Z. See [Berthelot–Ogus, 5.28]. This claim can be understood as a consequence of the existence of local retractions for infinitesimal thickenings in the smooth situation.

[25] "Un cristal possède deux propriétés caractéristiques: la rigidité, et la faculté de croître, dans un voisinage approprié." – Alexander Grothendieck, letter to John Tate.

close points.
Given a crystal in $\mathcal{O}_{Y/S}$-modules $\mathcal{M}$, we can consider the thickenings of Y given by the PD-envelopes of infinitesimal neighbourhoods of the diagonal of Y in $Y \times_S Y$. Starting with the simplicial scheme

$$\cdots \qquad Y \times_S Y \times_S Y \;\mathrel{\substack{\rightarrow\\ \rightarrow\\ \rightarrow\\ \leftarrow\\ \leftarrow}}\; Y \times_S Y \;\mathrel{\substack{\rightarrow\\ \rightarrow\\ \leftarrow}}\; Y,$$

we look at the various neighbourhoods of the diagonal. Let $\Delta^n_i(Y/S)$ denote the PD-envelope of the n-th order infinitesimal neighbourhood of the diagonal of Y in the $(i+1)$-fold fibred product of Y over S. These fit into the simplicial diagrams

$$\cdots \qquad \Delta^n_2(Y/S) \;\mathrel{\substack{\rightarrow\\ \rightarrow\\ \rightarrow\\ \leftarrow\\ \leftarrow}}\; \Delta^n_1(Y/S) \;\mathrel{\substack{\rightarrow\\ \rightarrow\\ \leftarrow}}\; Y.$$

Consider for the moment the case of first-order infinitesimal neighbourhoods, i.e. $n = 1$. Write $\pi_1, \pi_2 \colon \Delta^1_1(Y) \rightrightarrows Y$ for the two projection morphisms induced from $Y \times_S Y \rightrightarrows Y$. The condition for $\mathcal{M}$ to be a crystal implies that we must have an isomorphism $\alpha \colon \pi_1^* \mathcal{M} \xrightarrow{\sim} \pi_2^* \mathcal{M}$. This encapsulates infinitesimal parallel transport: given two points x, y of Y which are infinitesimally close to first order, i.e. a point $(x, y) \in \Delta^1_1(Y/S)$, we obtain an isomorphism $\eta_{x,y} \colon \mathcal{M}_x \xrightarrow{\sim} \mathcal{M}_y$. We also want these isomorphisms to be composable (i.e. we need transitivity of parallel transport). To obtain this, we bring in the arrows $\Delta^1_2(Y/S) \mathrel{\substack{\xrightarrow{\pi_{12}}\\ \xrightarrow{\pi_{13}}\\ \xrightarrow{\pi_{23}}}} \Delta^1_1(Y/S)$ induced by the projection morphisms from $Y \times_S Y \times_S Y$ to $Y \times_S Y$. The compatibility of the system of isomorphisms of the crystal $\mathcal{M}$ then implies that we have $\pi_{23}^*(\alpha) \circ \pi_{12}^*(\alpha) = \pi_{13}^*(\alpha)$. This cocycle condition means that we have descent data for $\mathcal{M}$ along $f \colon Y \to S$. In terms of the parallel transport, it means we have transitivity $\eta_{x,z} = \eta_{y,z} \circ \eta_{x,y}$; hence we can transport sections of $\mathcal{M}$ to first-order. This data of parallel transport for first-order infinitesimal neighbourhoods is the algebraic analogue of an Ehresmann connection, and is known as a **Grothendieck connection** (or Grothendieck 1-connection, to emphasise its first-order nature). If $f \colon Y \to S$ is smooth, this is equivalent to the data of a $\mathcal{O}_Y$-module $\mathcal{E}$ equipped Koszul connection, i.e. an $\mathcal{O}_S$-linear morphism $\nabla \colon \mathcal{E} \to \mathcal{E} \otimes_{\mathcal{O}_Y} \Omega^1_{Y/S}$ satisfying the Leibniz rule [Grothendieck. Crystals, Appendix].
We can also extend the notion of a Grothendieck connection to higher order neighbourhoods. A compatible system of n-th order Grothendieck connections (defined as above, including the cocycle condition, but using $\Delta^n_1(Y/S)$ instead of $\Delta^1_1(Y/S)$) is called a **(PD-)stratification**.
Continuing to assume smoothness of $f \colon Y \to S$, in characteristic 0, there is a three-way equivalence between (Zariski) sheaves of $\mathcal{O}_Y$-modules equipped with a stratification, (Zariski) sheaves of $\mathcal{O}_Y$-modules equipped with a *flat* Koszul connection, and crystals in $\mathcal{O}_{Y/S}$-modules [Berthelot–Ogus, Theorem 4.8]. Flatness of the connection corresponds to the curvature $F_\nabla = \nabla_1 \circ \nabla$ of ∇ being 0.[26] Flatness is automatic for crystals because we have parallel transport to higher order (i.e. natural transitivity formulæ$\eta_{x,z} = \eta_{y,z} \circ \eta_{x,z}$ valid at all orders), and the curvature of a connection precisely measures the holonomy of second-order parallel transport.
The situation is slightly more complicated in characteristic p. In this case we need to work with **hyper PD-stratifications**, signifying that instead of doing a compatible system of n-connections at all levels, we work directly with all levels at once by using the full divided power envelope of the diagonal of Y in $Y \times_S Y$ (which in a sense adds PD structures to all order infinitesimal neighbourhoods simultaneously).[27] These hyper PD-stratifications then match up with crystals on Y/S, so that the category of crystals of $\mathcal{O}_{Y/S}$-modules (with $\mathcal{O}_{Y/S}$-linear morphisms) is equivalent to that of $\mathcal{O}_Y$-modules equipped with a hyper PD-stratification (with morphisms the horizontal $\mathcal{O}_Y$-module homomorphisms, i.e. morphisms compatible with the stratification) [Berthelot–Ogus, Theorem 6.6].

[26] Recall that, given $\nabla \colon \mathcal{E} \to \mathcal{E} \otimes_{\mathcal{O}_Y} \Omega^1_{Y/S}$, we can define $\nabla_i \colon \mathcal{E} \otimes_{\mathcal{O}_Y} \Omega^i_{Y/S} \to \mathcal{E} \otimes_{\mathcal{O}_Y} \Omega^{i+1}_{Y/S}$ by the formula $\nabla_i(s \otimes \alpha) = s \otimes \mathrm{d}\alpha + (-1)^i \nabla s \otimes \alpha$. If $F_\nabla = 0$, these then define a complex, the de Rham complex of $\mathcal{E}$ with respect to ∇. We also have the formula $F_\nabla(\theta, \eta) = [\nabla_\theta, \nabla_\eta] - \nabla_{[\theta,\eta]}$.
[27] On the level of connections, this corresponds to $\mathcal{O}_Y$-modules equipped with a *quasi-nilpotent* flat Koszul connection.

5.3.2 The crystalline Poincaré lemma

We now look to the Poincaré lemma for solace. Assuming that $f: Y \to S$ is smooth, the idea is that, given a crystal $\mathcal{M}$ in $(Y/S)_{\mathrm{cris}}$, we can consider the associated $\mathcal{O}_Y$-module $\mathcal{E}_{\mathcal{M}} = \mathcal{M}_{(Y,Y,\mathrm{id})}$ (on the Zariski site of Y) with its flat connection ∇, and thus its de Rham complex $\mathcal{E}_{\mathcal{M}} \otimes_{\mathcal{O}_Y} \Omega^\bullet_{Y/S}$.

On the other hand, given the Zariski sheaves $\Omega^\bullet_{Y/S}$, we can turn them into crystals using a particular linearisation operation $\mathcal{L}$. Writing $\Delta_i(Y/S)$ for the formal scheme which is the PD-envelope of the infinitesimal neighbourhood (of all orders) of Y in the $(i+1)$-fold fibered product of Y over S, we have two projection morphisms $\pi_1, \pi_2: \Delta_1(Y/S) \rightrightarrows Y$. We then define $\mathcal{L}_Y\mathcal{F} = (\pi_1)_* \circ (\pi_2)^*\mathcal{F}$. Writing $\mathfrak{D}_i(Y/S) = \mathcal{O}_{\Delta_i(Y/S)}$, we can rewrite $\mathcal{L}_Y\mathcal{F} = \mathfrak{D}_1(Y/S) \otimes_{\mathcal{O}_Y} \mathcal{F}$, where the tensor product is taken with respect to the right $\mathcal{O}_Y$-module structure of $\mathfrak{D}_1(Y/S)$ (given by π_2), and the $\mathcal{O}_Y$-module structure of the tensor product is given by the left $\mathcal{O}_Y$-module structure of $\mathfrak{D}_1(Y/S)$ (given by π_1) [EGA IV$_4$, 16.7.1.2, 16.7.2.1]. We then say that a **hyper PD-differential operator** between $\mathcal{O}_Y$-modules $\mathcal{F} \to \mathcal{G}$ is an $\mathcal{O}_Y$-linear homomorphism $\mathfrak{D}_1(Y/S) \otimes_{\mathcal{O}_Y} \mathcal{F} \to \mathcal{G}$. Through the linearisation functor, these precisely correspond to hyper PD-stratified $\mathcal{O}_Y$-modules. Indeed, given any (Zariski) $\mathcal{O}_Y$-module $\mathcal{F}$, $\mathcal{L}_Y\mathcal{F}$ is naturally equipped with a hyper PD-stratification, and $\mathcal{L}_Y$ turns hyper PD-differential operators into $\mathcal{O}_Y$-linear maps, and is exact if $f: Y \to S$ is smooth [Berthelot–Ogus, Construction 6.9]. Assuming then the smoothness of f, exactness of $\mathcal{L}_Y$ means that in our situation we get a complex of crystals $\mathcal{L}_Y(\mathcal{E}_{\mathcal{M}} \otimes_{\mathcal{O}_Y} \Omega^\bullet_{Y/S}) \cong \mathcal{M} \otimes_{\mathcal{O}_{Y/S}} \mathcal{L}_Y(\Omega^\bullet_{Y/S})$.[28]

We can then state the crystalline version of the Poincaré lemma:

Theorem 5.1 — The complex $\mathcal{M} \otimes_{\mathcal{O}_{Y/S}} \mathcal{L}_Y(\Omega^\bullet_{Y/S})$ is a resolution of $\mathcal{M}$ in the category of sheaves of $\mathcal{O}_{Y/S}$-modules on $\mathrm{Cris}(Y/S)$. ◇

Proof: See [Berthelot. H$_{\mathrm{cris}}$, Théorème V.2.1.1]. The idea is that $\mathcal{L}_Y(\Omega^i_{Y/S})$ can be explicitly described: by dévissage and smoothness of f, we can assume that we have sections $y_1, \ldots, y_d$ of $\mathcal{O}_Y$ such that $\mathrm{d}y_1, \ldots, \mathrm{d}y_d$ form a basis of $\Omega^1_{Y/S}$. Writing ξ_i for the image of $1 \otimes y_i - y_i \otimes 1$ in $\mathfrak{D}_1(Y/S)$, we have the description

$$\mathcal{L}_Y(\Omega^i_{Y/S}) = \mathcal{O}_Y\langle \xi_1, \ldots, \xi_d\rangle \otimes_{\mathcal{O}_Y[\xi_1,\ldots,\xi_d]} \Omega^i_{\mathcal{O}_Y[\xi_1,\ldots,\xi_d]/\mathcal{O}_Y}.$$

This reduces the situation to the following local statement: for any ring R, the complex

$$R\langle t_1, \ldots, t_n\rangle \otimes_{R[t_1,\ldots,t_n]} \Omega^\bullet_{R[t_1,\ldots,t_n]/R},$$

which is the de Rham complex of $R[t_1, \ldots, t_n]/R$ with coefficients in the $R[t_1, \ldots, t_n]$-algebra of divided power polynomials $R\langle t_1, \ldots, t_n\rangle$ (and connection given by $\nabla(t_i^{[k]}) = t_i^{[k-1]} \otimes \mathrm{d}t_i$), is a resolution of R in the abelian category of R-modules [Berthelot. H$_{\mathrm{cris}}$, Lemme V.2.1.2]. □

5.3.3 Crystalline cohomology and de Rham cohomology

We can now leverage the Poincaré lemma to relate the crystalline and de Rham cohomologies. The idea is to take inspiration from **Alexander–Spanier cohomology**.

Given a topological space X, and a (discrete) abelian group A, we define the **Čech–Alexander complex** $\check{\mathrm{C}}\mathrm{A}^\bullet(X, A)$ by letting $\check{\mathrm{C}}\mathrm{A}^i(X, A)$ be the sheaf associated with the presheaf of continuous functions

$$U \mapsto \{f: U^{i+1} \to A \mid f \text{ vanishes in a neighbourhood of the diagonal of } U \text{ in } U^{i+1}\}.$$

In other words, $\check{\mathrm{C}}\mathrm{A}^i(X, A)$ is the sheaf of germs of A-valued functions on the diagonal of X in X^{i+1}.

The differentials are given by the customary Čech-style alternating sum formulæ, e.g. for $f: U \to A$, $df(x, y) = f(x) - f(y)$, and for $g: U^2 \to A$, $dg(x, y, z) = g(x, y) - g(x, z) + g(y, z)$. The cohomology of this complex is the Alexander-Spanier cohomology of X with coefficients in A, and it computes the usual singular cohomology groups in many circumstances, for instance if X is (homotopy equivalent to) a CW-complex.

In our context, we would like to replace this complex by a complex of functions on infinitesimal neighbourhoods of the diagonal. So, given as before $(\underline{S}, \mathcal{I}, \gamma)$, and $f: Y \to S$ with γ extending to Y, we define the cosimplicial

[28] This isomorphism requires that $\mathcal{M}$ be a crystal, and not just an arbitrary sheaf of $\mathcal{O}_{Y/S}$-modules [Berthelot. H$_{\mathrm{cris}}$, Proposition IV.3.1.4].

object $\check{\mathrm{C}}\mathrm{A}^i_Y(\mathcal{F}) = \mathcal{F}|_{(Y,\Delta_i(Y/S),\iota_{\Delta_i(Y/S)})}$. When $\mathcal{F}$ is a sheaf of abelian groups, we can again turn this cosimplicial object into a complex of sheaves on Y_{Zar} by taking differentials to be alternating sums of face maps, yielding the crystalline Čech–Alexander complex of $\mathcal{F}$.

The usefulness of the Čech–Alexander complex stems from its relationship with $\mathcal{L}_Y$: given any (Zariski) $\mathcal{O}_Y$-module $\mathcal{F}$, $\check{\mathrm{C}}\mathrm{A}^\bullet(\mathcal{L}_Y\mathcal{F})$ is quasi-isomorphic to $\mathcal{F}$ concentrated in degree 0 [Berthelot. $\mathrm{H}_{\mathrm{cris}}$, Proposition V.2.2.2]. This is essentially because $\mathcal{L}_Y$ shifts along the Čech–Alexander complex by one. To see this, we use that $\mathfrak{D}_i(Y/S) \otimes_{\mathcal{O}_Y} \mathfrak{D}_j(Y/S) = \mathfrak{D}_{i+j}(Y/S)$ [Berthelot. $\mathrm{H}_{\mathrm{cris}}$, Lemme II.1.3.4]. Then, from the definitions of $\mathcal{L}_Y$ and $\check{\mathrm{C}}\mathrm{A}^\bullet_Y$, we obtain that $\check{\mathrm{C}}\mathrm{A}^i_Y(\mathcal{L}_Y\mathcal{O}_Y) = \check{\mathrm{C}}\mathrm{A}^{i+1}_Y(\mathcal{O}_Y)$ for $i \geqslant 0$, and $\check{\mathrm{C}}\mathrm{A}^i_Y(\mathcal{L}_Y\mathcal{O}_Y) = 0$ for $i < 0$. Moreover, the cosimplicial object $0 \to \check{\mathrm{C}}\mathrm{A}^0_Y(\mathcal{O}_Y) \to \check{\mathrm{C}}\mathrm{A}^\bullet_Y(\mathcal{L}_Y\mathcal{O}_Y)$ has an extra degeneracy, and so is homotopic to zero [Berthelot. $\mathrm{H}_{\mathrm{cris}}$, Lemme V.2.2.1]. This works the same for other coefficients, showing that $\check{\mathrm{C}}\mathrm{A}^\bullet_Y(\mathcal{L}_Y\mathcal{F})$ is quasi-isomorphic to $\check{\mathrm{C}}\mathrm{A}^0_Y(\mathcal{F}) = \mathcal{F}$.
Now, in this context, the Čech–Alexander complex computes crystalline cohomology:

Theorem 5.2 — Let $\mathcal{R}$ be a sheaf of rings on $\mathrm{Cris}(Y/S)$, $\mathcal{F}^\bullet$ a bounded below complex of $\mathcal{R}$-modules. Assuming $f: Y \to S$ is smooth, there is a canonical isomorphism of complexes of $\mathrm{u}_*(\mathcal{R})$-modules in the derived category of Zariski sheaves on Y:

$$\mathbf{R}\mathrm{u}_*\mathcal{F}^\bullet \cong \mathrm{Tot}\,\check{\mathrm{C}}\mathrm{A}^\bullet_Y(\mathcal{F}^\bullet). \qquad \diamond$$

Proof: See [Berthelot. $\mathrm{H}_{\mathrm{cris}}$, Théorème V.1.2.5]. The smoothness hypothesis allows us to work with the crystalline topos, instead of the "hyper PD-stratifying topos". □

Putting it all together now, we have isomorphisms:

$$\begin{aligned}
\mathrm{H}^i_{\mathrm{cris}}(Y,\mathcal{M}) &\cong \mathbf{H}^i_{\mathrm{cris}}\left(Y, \mathcal{M} \otimes_{\mathcal{O}_{Y/S}} \mathcal{L}_Y\left(\Omega^\bullet_{Y/S}\right)\right) \cong \mathbf{H}^i_{\mathrm{cris}}\left(Y, \mathcal{L}_Y\left(\mathcal{E}_{\mathcal{M}} \otimes_{\mathcal{O}_Y} \Omega^\bullet_{Y/S}\right)\right) \\
&\cong \mathbf{H}^i_{\mathrm{Zar}}\left(Y, \mathrm{Tot}\,\check{\mathrm{C}}\mathrm{A}^\bullet_Y\left(\mathcal{L}_Y\left(\mathcal{E}_{\mathcal{M}} \otimes_{\mathcal{O}_Y} \Omega^\bullet_{Y/S}\right)\right)\right) \quad \text{by Theorem 5.2} \\
&\cong \mathbf{H}^i_{\mathrm{Zar}}\left(Y, \mathcal{E}_{\mathcal{M}} \otimes_{\mathcal{O}_Y} \Omega^\bullet_{Y/S}\right) \quad \text{as } \mathcal{F} \xrightarrow{\text{q.iso}} \check{\mathrm{C}}\mathrm{A}^\bullet(\mathcal{L}_Y\mathcal{F}) \\
&= \mathrm{H}^i_{\mathrm{dR}}(Y/S, \mathcal{E}_{\mathcal{M}}).
\end{aligned}$$

This method can also be applied to compute the crystalline cohomology in other cases. The idea is to embed an arbitrary $f: Y \to S$ into a smooth scheme Z through a closed immersion $j: Y \to Z$, and consider the PD-envelope of Y in Z, $\mathrm{D}_Y(Z)$. In this case, for a $\mathcal{O}_Y$-module $\mathcal{E}$ equipped with a hyper PD-stratification, and $\mathcal{M}$ the corresponding crystal on $\mathrm{Cris}(Y/S)$ obtained by restriction, we have [Berthelot. $\mathrm{H}_{\mathrm{cris}}$, Théorème V.2.3.2]:

$$\mathbf{R}(\mathrm{u}_{Y/S})_*\mathcal{M} \cong \mathcal{O}_{\mathrm{D}_Y(Z)} \otimes_{\mathcal{O}_Z} \mathcal{E}_{\mathcal{M}} \otimes_{\mathcal{O}_Z} \Omega^\bullet_{Z/S}$$

in the derived category of sheaves of $f^{-1}(\mathcal{O}_S)$-modules in Y_{Zar}.

Now, we just saw that the crystalline cohomology and the de Rham cohomology of the special fibre, over κ, are isomorphic. We want more! Let's involve $\mathrm{H}^i_{\mathrm{cris}}(Y/W, \mathcal{O}_{Y/W})$ and the de Rham cohomology of the generic fibre $\mathrm{H}^i_{\mathrm{dR}}(X/K)$. We only needed the complication of divided powers to account for non-invertibility of p (which led to problems with expressions of the form $x^p/p!$), so it stands to reason that one might have

$$\mathrm{H}^i_{\mathrm{cris}}(Y/W, \mathcal{O}_{Y/W}) \otimes_W K \cong \mathrm{H}^i_{\mathrm{dR}}(X/K).$$

This is in fact the case, although it is far from obvious. The previous considerations involving the crystalline Poincaré lemma and the Čech–Alexander complex show that we in fact have an isomorphism at the integral level if Y has a smooth lift $\mathcal{X}$ to W (and not just a lift to $\mathcal{O}_K$) [Berthelot–Ogus, Theorem 7.23, 7.26.3]

$$\mathrm{H}^i_{\mathrm{cris}}(Y/W, \mathcal{O}_{Y/W}) \xrightarrow{\sim} \mathrm{H}^i_{\mathrm{dR}}(\mathcal{X}/W).$$

In general however, we do really need to tensor to K.

Theorem 5.3 (Berthelot, Ogus) — Let X/K be a smooth scheme, with smooth integral model $\mathcal{X}/\mathcal{O}_K$ and special fibre $Y = \mathcal{X}_\kappa$. Then:

$$\mathrm{H}^i_{\mathrm{cris}}(Y/W, \mathcal{O}_{Y/W}) \otimes_W K \xrightarrow{\sim} \mathrm{H}^i_{\mathrm{dR}}(X/K). \qquad \diamond\square$$

We cannot expect to strengthen this isomorphism to the integral level to obtain an isomorphism $\mathrm{H}^i_{\mathrm{cris}}(Y/W, \mathcal{O}_{Y/W}) \to \mathrm{H}^i_{\mathrm{dR}}(\mathcal{X}/W)$, unless $e(K) \leqslant p-1$ (so that the maximal ideal $\mathfrak{m}$ of $\mathcal{O}_K$ admits a PD-structure) [Illusie. $\mathrm{H}_{\mathrm{cris}}$, 1.3 (b), 2.3].

We can also use this notion of crystals to re-interpret the Gauss–Manin connection from section 1.2.1. The idea is to use a six operations formalism. Given a smooth morphism $g\colon Y_1 \to Y_2$ of schemes over $(S, \mathcal{I}, \gamma)$ (with γ extending to Y_1 and Y_2, moreover Y_2 is assumed smooth over S), and a crystal $\mathcal{M}$ on Y_1, we can consider $\mathrm{R}g_*\mathcal{M}$ as a crystal, with the complication now that we have to deal with derived categories; so $\mathrm{R}g_*\mathcal{M}$ is a crystal with values in the fibred category whose fibre above an object $(U, T, \delta) \in \mathrm{Cris}(Y_2/S)$ is the derived category of the category of $\mathcal{O}_T$-modules. For this to hold we need to assume a few conditions: $\mathcal{M}$ is a crystal of quasi-coherent flat modules, Y_2 quasi-compact, g is quasi-compact and quasi-separated [Berthelot. $\mathrm{H}_{\mathrm{cris}}$, Théorème V.3.6.1]. With these conditions assumed, we obtain a flat quasi-nilpotent connection on each $\mathrm{R}^i g_*\mathcal{E}$; the derived nature does not pose any complications because the projection morphisms $\Delta_1(Y_2/S) \rightrightarrows Y_2$ are flat [Berthelot. $\mathrm{H}_{\mathrm{cris}}$, Corollaire V.3.6.3].
Now, we have a quasi-isomorphism $\mathrm{R}g_*\mathcal{M}|_{(Y_2,Y_2,\mathrm{id})} \cong \mathbf{R}g_*\left(\mathcal{E}_{\mathcal{M}} \otimes_{\mathcal{O}_{Y_1}} \Omega^\bullet_{Y_1/Y_2}\right)$ (where on the left g_* is pushforward of crystalline sheaves, and on the right of Zariski sheaves). This allows us to endow each $\mathbf{R}^i g_*\left(\mathcal{E}_{\mathcal{M}} \otimes_{\mathcal{O}_{Y_1}} \Omega^\bullet_{Y_1/Y_2}\right)$ with a hyper PD-stratification over S, and so a flat quasi-nilpotent connection over S, which is exactly the Gauss–Manin connection on the relative algebraic de Rham cohomology [Berthelot. $\mathrm{H}_{\mathrm{cris}}$, Proposition V.3.6.4].

Finally, we note that crystalline cohomology is indeed a Weil cohomology theory, as was shown by Berthelot. One has cycle class maps [Berthelot. $\mathrm{H}_{\mathrm{cris}}$, Théorème VI.3.3.5], a Künneth formula [Berthelot. $\mathrm{H}_{\mathrm{cris}}$, Théorème V.4.2.1], Poincaré duality [Berthelot. $\mathrm{H}_{\mathrm{cris}}$, Théorème VII.2.1.3], and so on.

5.3.4 The de Rham–Witt complex

How do we compute crystalline cohomology? We would like to explicitly compute the functor $\mathrm{R}u_*$. The Čech–Alexander complex provided us with one way of doing so, but it is not particularly explicit.
We saw that crystalline cohomology and de Rham cohomology of the special fibre (with coefficients in the residue field κ) are isomorphic. We can compute the crystalline cohomology over κ by using the de Rham complex $\Omega^\bullet_{X/\kappa}$; is there a complex over W that does a similar job?
The idea goes back to Serre, who proposed the cohomology theory $Y \mapsto \mathrm{H}^n_{\mathrm{Zar}}(Y, \mathcal{W}\mathcal{O}_Y)$ as a way to tackle the Weil conjectures. This is however not a Weil cohomology theory, as, just like coherent cohomology, the cohomology groups vanish for $n > \dim Y$. However, it gives us a starting point. We would like to define a complex $\mathcal{W}_\star\Omega^\bullet_Y$ that extends $\mathcal{W}_\star\mathcal{O}_Y$ in degree 0. An obvious first choice would be $\Omega^\bullet_Y \otimes_W \mathcal{W}_\star\mathcal{O}_Y$, but this is not quite the right object; we want to preserve the universal property of the Witt vectors: it is the universal lift to characteristic 0 that also lifts Frobenius.

Definition 5.4 — Let κ be a perfect field of characteristic p. A **Witt complex** is a projective system $E^\bullet_\star = (E^\bullet)_\star$ of (graded-)commutative differential graded W-algebras, with grading $(-)^\bullet$ and projective indexing $(-)_\star$, equipped with homomorphisms of pro-objects:

- of rings $\lambda\colon \mathcal{W}_\star(\kappa) \to E^0_\star$,
- of rings $F\colon E^\bullet_\star \to E^\bullet_{\star-1}$,
- of $E^\bullet_\star$-modules $V\colon FE^\bullet_\star \to E^\bullet_{\star-1}$,

with additional relations:

- $F\lambda = \lambda F$, $V\lambda = \lambda V$,
- $FV = VF = p$,
- $Fd\lambda([a]) = \lambda([a]^{p-1})d\lambda([a])$ (where $[-]$ indicates the Teichmüller lift $\kappa \to \mathcal{W}_\star(\kappa)$),[29]
- $V(F(x)y) = xVy$ (the projection formula),
- $FdV = d$. $\diamond$

We can visualise a Witt complex $E^\bullet_\star$ by the diagram:

$$\begin{tikzcd}
 & \vdots \arrow[d] \arrow[d, bend left, "F"] & \vdots \arrow[d] \arrow[d, bend left, "F"] & \vdots \arrow[d] \arrow[d, bend left, "F"] & \vdots \arrow[d] \arrow[d, bend left, "F"] & \iddots \\
 & \mathcal{W}_3(\kappa) \arrow[r, "\lambda"] \arrow[d] \arrow[d, bend left, "F"] \arrow[u, bend left, "V"] & E^0_3 \arrow[r, "d"] \arrow[d] \arrow[d, bend left, "F"] \arrow[u, bend left, "V"] & E^1_3 \arrow[r, "d"] \arrow[d] \arrow[d, bend left, "F"] \arrow[u, bend left, "V"] & E^2_3 \arrow[r, "d"] \arrow[d] \arrow[d, bend left, "F"] \arrow[u, bend left, "V"] & \cdots \\
 & \mathcal{W}_2(\kappa) \arrow[r, "\lambda"] \arrow[d] \arrow[d, bend left, "F"] \arrow[u, bend left, "V"] & E^0_2 \arrow[r, "d"] \arrow[d] \arrow[d, bend left, "F"] \arrow[u, bend left, "V"] & E^1_2 \arrow[r, "d"] \arrow[d] \arrow[d, bend left, "F"] \arrow[u, bend left, "V"] & E^2_2 \arrow[r, "d"] \arrow[d] \arrow[d, bend left, "F"] \arrow[u, bend left, "V"] & \cdots \\
\kappa \arrow[r, "{[-]}"] \arrow[ru, bend left, "{[-]}"] \arrow[ruu, bend left, "{[-]}"] & \mathcal{W}_1(\kappa) \arrow[r, "\lambda"] \arrow[u, bend left, "V"] & E^0_1 \arrow[r, "d"] \arrow[u, bend left, "V"] & E^1_1 \arrow[r, "d"] \arrow[u, bend left, "V"] & E^2_1 \arrow[r, "d"] \arrow[u, bend left, "V"] & \cdots
\end{tikzcd}$$

The underlying principle of this definition is that $\mathcal{W}_\star\Omega^\bullet_Y$ should be the universal Witt complex over Y. We can bootstrap the construction, starting with $\mathcal{W}_1\Omega^\bullet_Y = \Omega^\bullet_Y$ and $\mathcal{W}_\star\Omega^0_Y = \mathcal{W}_\star\mathcal{O}_Y$, to then inductively construct $\mathcal{W}_n\Omega^\bullet_Y$. Essentially, the idea is to start with $\Omega^\bullet_{\mathcal{W}_\star(\mathcal{O}_Y)}$, and force the existence of the required Frobenius and Verschiebung morphisms by imposing certain relations. See [Illusie. $\mathrm{W}_\star\Omega^\bullet$, Théorème I.1.3] for the details of the contruction of $\mathcal{W}_\star\Omega^\bullet_Y$, and the proof that it indeed yields the universal Witt complex over Y.

Finally, we want to know that this indeed computes the crystalline cohomology of Y when Y is smooth. We have:

Theorem 5.5 (Bloch, Illusie) — Let S be a perfect scheme of characteristic p, and $f\colon Y \to S$ a scheme of finite type over S. There is a natural morphism in the derived category of (Zariski) sheaves of $f^*(\mathcal{W}_n\mathcal{O}_S)$-modules on Y

$$\mathrm{R}(\mathrm{u}_{Y/\mathcal{W}_n(S)})_*\mathcal{O}_{Y/\mathcal{W}_n(S)} \longrightarrow \mathcal{W}_n\Omega^\bullet_Y,$$

which is a quasi-isomorphism if $f\colon Y \to S$ is smooth. $\diamond$

Proof: See [Illusie. $\mathrm{W}_\star\Omega^\bullet$, Proposition II.1.2] and [Illusie. $\mathrm{W}_\star\Omega^\bullet$, Théorème II.1.4]. $\square$

Recall that u is the forgetful functor from the crystalline site to the Zariski site. We saw in section 5.3.1 that we could compute the crystalline cohomology of Y (a smooth scheme of finite type over κ) as the (Zariski) hypercohomology of the sheaf $\mathrm{u}_*\mathcal{O}_{Y/W}$. The above theorem then guarantees that we can also compute the crystalline cohomology of Y as the hypercohomology of the de Rham–Witt complex of Y:

$$\mathrm{H}^i_{\mathrm{cris}}\left(Y/W, \mathcal{O}_{Y/W}\right) = \mathbf{H}^i_{\mathrm{Zar}}(Y, \mathcal{W}\Omega^\bullet_Y).$$

5.4 The definition of $\mathbf{B}^+_{\mathrm{cris}}$

As noted in section 3, we lost the Frobenius endomorphism in going from $\mathbf{A}_{\mathrm{inf}}$ to $\mathbf{B}^+_{\mathrm{dR}}$. The problem was that completing at $\ker(\theta)$ was too rough. Instead, let's go back to $\mathbf{A}_{\mathrm{inf}}$, which had the property of being the

[29] This is a "divided-by-p" Frobenius morphism on the level of differentials, as we have taken the factor of p out of the equation $\mathrm{d}x^p = px^{p-1}\mathrm{d}x$.

universal p-adic infinitesimal $\mathcal{O}_K$-thickening of $\mathcal{O}_{\mathbb{C}_p}$, and add divided powers to obtain a universal PD-thickening.

More specifically, starting with $\mathbf{A}_{\mathrm{inf}}$, instead of inverting p (thus obtaining all elements of the form ${x^n}/{m}$ for $x \in \ker(\theta)$), we should instead take the PD-envelope of $\ker(\theta)$ in $\mathbf{A}_{\mathrm{inf}}$; that is, adjoin elements of the form ${x^n}/{n!}$ to $\mathbf{A}_{\mathrm{inf}}$, for $x \in \ker(\theta)$. Taking the generator $\xi = \tilde{p} - p$ of $\ker(\theta)$, it suffices to adjoin elements of the form ${\xi^n}/{n!}$. This defines the ring $\mathbf{A}^0_{\mathrm{cris}}$; we are not done yet however as this ring has the misfortune of not being p-adically complete. Performing the p-adic completion of $\mathbf{A}^0_{\mathrm{cris}}$, we obtain $\mathbf{A}_{\mathrm{cris}}$. It can be shown, although it is not obvious, that $\mathbf{A}^0_{\mathrm{cris}}$ is at least p-adically separated, so that the natural map $\mathbf{A}^0_{\mathrm{cris}} \to \mathbf{A}_{\mathrm{cris}}$ is injective. We write again $\theta\colon \mathbf{A}_{\mathrm{cris}} \to \mathcal{O}_{\mathbb{C}_p}$ for the morphism induced by $\theta\colon \mathbf{A}_{\mathrm{inf}} \to \mathcal{O}_{\mathbb{C}_p}$.
The upshot is that, just as $\mathbf{A}_{\mathrm{inf}}$ was the universal p-adic infinitesimal $\mathcal{O}_K$-thickening of $\mathcal{O}_{\mathbb{C}_p}$, $\mathbf{A}_{\mathrm{cris}}$ is the universal p-adic infinitesimal $\mathcal{O}_{K_0}$-PD-thickening of $\mathcal{O}_{\mathbb{C}_p}$.
The ring $\mathbf{A}_{\mathrm{cris}}$ contains t, since by writing $1 - \tilde{\varepsilon} = a\xi$, we have:

$$\mathrm{t} = -\sum_{n \geqslant 1} \frac{(1-\tilde{\varepsilon})^n}{n} = -\sum_{n \geqslant 1} (n-1)!a^n \frac{\xi^n}{n!}.$$

We can then define $\mathbf{B}^+_{\mathrm{cris}} = \mathbf{A}_{\mathrm{cris}}[1/p]$, $\mathbf{B}_{\mathrm{cris}} = \mathbf{B}^+_{\mathrm{cris}}[1/t]$. This is Fontaine's **ring of crystalline periods**.[30]
It is possible to show that $\mathbf{B}^+_{\mathrm{cris}}$ embeds into $\mathbf{B}^+_{\mathrm{dR}}$, with $\mathbf{A}_{\mathrm{cris}}$ given the p-adic topology.[31] This then means we can describe $\mathbf{A}_{\mathrm{cris}}$ and $\mathbf{B}^+_{\mathrm{cris}}$ as the following subrings of $\mathbf{B}^+_{\mathrm{dR}}$:

$$\mathbf{A}_{\mathrm{cris}} = \left\{ \sum_{n \geqslant 0} a_n \frac{\xi^n}{n!} \in \mathbf{B}^+_{\mathrm{dR}} \;\middle|\; a_n \in \mathbf{A}_{\mathrm{inf}},\ (a_n) \to 0 \right\},$$
$$\mathbf{B}^+_{\mathrm{cris}} = \left\{ \sum_{n \geqslant 0} a_n \frac{\xi^n}{n!} \in \mathbf{B}^+_{\mathrm{dR}} \;\middle|\; a_n \in \mathbf{A}_{\mathrm{inf}}[1/p],\ (a_n) \to 0 \right\}.$$

From the description of $\mathbf{A}_{\mathrm{cris}}$ as a universal PD-thickening, we expect to be able to write $\mathbf{A}_{\mathrm{cris}}$ in terms of crystalline cohomology, just as we used the infinitesimal topos to give another definition of $\mathbf{A}_{\mathrm{inf}}$. Mirroring section 3, but using the crystalline topos of section 5.3, given a scheme $Z/\mathcal{W}_n(\kappa)$, we define $\mathbb{O}^{\mathrm{cris}}_n(Z) = \mathrm{H}^0_{\mathrm{cris}}\left(Z/\mathcal{W}_n(\kappa), \mathbb{O}_{Z/\mathcal{W}_n(\kappa)}\right)$ (where $\kappa = \mathcal{O}_K/\mathfrak{m}_K$). We then have $\mathbf{A}_{\mathrm{cris}} = \varprojlim_n \mathbb{O}^{\mathrm{cris}}_n(\mathcal{O}_{\overline{K}}/(p^n))$. Once again, doing it in levels and then taking an inverse limit like this allows us to recover the topological structure from algebraic objects.

5.5 The fundamental exact sequence of $\mathbf{B}_{\mathrm{cris}}$

Finally, we want to consider all structures available on $\mathbf{B}_{\mathrm{cris}}$. The Galois action is inherited from that of $\mathbf{A}_{\mathrm{inf}}$, with the same formula as in section 3. We know that $\mathbf{A}_{\mathrm{inf}} \subset \mathbf{A}_{\mathrm{cris}}$, with $(\mathbf{A}_{\mathrm{inf}})^{\mathrm{G}_K} = W = \mathcal{O}_{K_0}$ as noted previously, so that certainly $\mathcal{O}_{K_0} \subseteq (\mathbf{A}_{\mathrm{cris}})^{\mathrm{G}_K}$ and $K_0 \subseteq \left(\mathbf{B}^+_{\mathrm{cris}}\right)^{\mathrm{G}_K}$. In fact these are all the invariants there are; moreover the natural Galois-equivariant homomorphism $\mathbf{B}_{\mathrm{cris}} \otimes_{K_0} K \to \mathbf{B}_{\mathrm{dR}}$ is injective. For a proof, see for instance [Brinon–Conrad, Theorem 9.1.5]. The Hodge filtration on $\mathbf{B}_{\mathrm{dR}}$ then pulls back to a filtration on $\mathbf{B}_{\mathrm{cris}}$.
Now, the whole point of the definition of $\mathbf{B}^+_{\mathrm{cris}}$ was to obtain a Frobenius endomorphism. We have:

$$\varphi(\xi) = \varphi(\tilde{p} - p) = \tilde{p}^p - p = (\xi + p)^p - p,$$

whence $\varphi(\xi) = \xi^p + pa$ for some $a \in \mathbf{A}_{\mathrm{inf}}$. We can rewrite this, using divided powers, as $\varphi(\xi) = p\left((p-1)!\xi^{[p]} + a\right)$, so that $\varphi(\xi^n) = p^n\left(a + (p-1)!\xi^{[p]}\right)^n$. As $\mathrm{v}_p(p^n) \geqslant \mathrm{v}_p(n!)$, this shows that $\mathbf{A}^0_{\mathrm{cris}}$ is stable under φ. Therefore φ extends to both $\mathbf{A}_{\mathrm{cris}}$ and $\mathbf{B}^+_{\mathrm{cris}}$. We also have that $\varphi(\mathrm{t}) = p\mathrm{t}$, as $\varphi(\log(\tilde{\varepsilon})) = \log(\tilde{\varepsilon}^p) = p\log(\tilde{\varepsilon}) = p\mathrm{t}$, as can be checked directly through the convergent series defining $\log(\tilde{\varepsilon})$.

[30] Another possibility, more in line with the definition of $\mathbf{B}^+_{\mathrm{dR}}$, would be to take the PD-completion of $\mathbf{A}_{\mathrm{cris}}$ (mirroring the completion step in going from $\mathbf{A}_{\mathrm{inf}}$ to $\mathbf{B}^+_{\mathrm{dR}}$); that is, we would consider $\left(\varprojlim_n \mathbf{A}_{\mathrm{cris}}/\ker(\theta)^{[n]}\right)[1/p]$. This is not particularly helpful, however, as this ring loses its Frobenius operator, once again because $\varphi(\xi) \notin \ker(\theta)$.
[31] The p-adic topology is the natural topology to put on $\mathbf{A}_{\mathrm{cris}}$; we only needed to modify the topology for $\mathbf{B}_{\mathrm{dR}}$ because of excessive ramification, which we are not allowing in the crystalline situation.

Hence φ also extends to $\mathbf{B}_{\mathrm{cris}}$. In particular, φ is a σ-semilinear operator on the K_0-algebra $\mathbf{B}_{\mathrm{cris}}$. Note however that φ does not commute with the Hodge filtration on $\mathbf{B}_{\mathrm{cris}}$, once again because $\varphi(\xi) \notin \ker(\theta)$.
Continuing to build on the description of $\mathbf{A}_{\mathrm{inf}}$ from section 3, we would like to better understand this Frobenius action on $\mathbf{B}_{\mathrm{cris}}$, and interaction with the Hodge filtration. We saw that $(\mathbf{A}_{\mathrm{inf}})^{\varphi^r} = \mathbb{Z}_{p^r}$. The situation for crystalline periods is more intricate, and captured by the following fundamental exact sequences.

Theorem 5.6 — Let $n \in \mathbb{N}$ be any natural number. The following sequence of $\mathbb{Q}_p[\mathrm{G}_K]$-modules is exact:

$$0 \longrightarrow \mathbb{Q}_p \cdot \mathrm{t}^n \longrightarrow (\mathbf{B}^+_{\mathrm{cris}})^{\varphi = p^n} \longrightarrow \mathbf{B}^+_{\mathrm{dR}}/(\mathrm{t}^n) \longrightarrow 0. \qquad \diamond\square$$

Corollary 5.7 — Let m be an arbitrary integer. We have that $F^m(\mathbf{B}_{\mathrm{cris}})^{\varphi = p^m} = \mathbb{Q}_p \cdot \mathrm{t}^m$. Moreover, the following sequences of $\mathbb{Q}_p[\mathrm{G}_K]$-modules are exact:

$$0 \longrightarrow \mathbb{Q}_p \cdot \mathrm{t}^m \longrightarrow (\mathbf{B}_{\mathrm{cris}})^{\varphi = p^m} \longrightarrow \mathbf{B}_{\mathrm{dR}}/\mathrm{t}^m\mathbf{B}^+_{\mathrm{dR}} \longrightarrow 0,$$

$$0 \longrightarrow \mathbb{Q}_p \cdot \mathrm{t}^m \longrightarrow F^m\mathbf{B}_{\mathrm{cris}} \xrightarrow{\varphi - p^m} \mathbf{B}_{\mathrm{cris}} \longrightarrow 0. \qquad \diamond$$

Proof: By twisting by t^{-m}, we may assume $m = 0$. The first exact sequence is obtained by taking the exact sequence of Theorem 5.6, then dividing by t^n to obtain

$$0 \longrightarrow \mathbb{Q}_p \longrightarrow \left(\mathrm{t}^{-n}\mathbf{B}^+_{\mathrm{cris}}\right)^{\varphi=1} \longrightarrow \mathrm{t}^{-n}\mathbf{B}^+_{\mathrm{dR}}/\mathbf{B}^+_{\mathrm{dR}} \longrightarrow 0,$$

before finally taking a direct limit over n.
To see that $F^0(\mathbf{B}_{\mathrm{cris}})^{\varphi=1} = \mathbb{Q}_p$, we use the first exact sequence. It tells us that

$$\mathbb{Q}_p = \ker\left((\varphi - 1)\colon \mathbf{B}_{\mathrm{cris}} \to \mathbf{B}_{\mathrm{dR}}/\mathbf{B}^+_{\mathrm{dR}}\right) = F^0(\mathbf{B}_{\mathrm{cris}})^{\varphi=1}.$$

This then implies exactness in the middle of the second sequence. Injectivity is clear, so it remains to show that $(\varphi - 1)\colon F^0\mathbf{B}_{\mathrm{cris}} \to \mathbf{B}_{\mathrm{cris}}$ is surjective. This is more difficult; see for instance [Fontaine. $\mathbf{B}_{\mathrm{dR}}$, 5.3.6]. $\square$

It is perhaps more natural to view this in terms of double complexes. The exactness of the first sequence in Corollary 5.7 (for $m = 0$) is precisely the exactness of the total complex of the following double complex:

$$\begin{tikzcd}
\mathbb{Q}_p \arrow[r] \arrow[d] & \mathbf{B}_{\mathrm{cris}} \arrow[r, "\varphi-1"] \arrow[d] & \mathbf{B}_{\mathrm{cris}} \\
\mathbf{B}^+_{\mathrm{dR}} \arrow[r] & \mathbf{B}_{\mathrm{dR}} &
\end{tikzcd}$$

6 The crystalline comparison theorem

We come back now to the setting of comparison theorems, with $\mathbf{B}_{\mathrm{cris}}$ in hand. To that effect, let X/K be a smooth proper variety, and assume now that X has *good reduction*: X extends to a smooth scheme $\mathcal{X}/\mathcal{O}_K$, with special fibre Y/κ. Grothendieck's hope was that one could do much better in this situation than what we managed to achieve so far with $\mathbf{B}_{\mathrm{dR}}$: he hoped we could recover the action of G_K purely from the other data at hand, i.e. the filtration and the Frobenius endomorphism. This was inspired from the situation of abelian varieties, and more generally of p-divisible groups, where one has the notion of Dieudonné modules.
Specifically, Grothendieck hoped that one could recover the p-adic étale cohomology $\mathrm{H}^n_{\text{ét}}\left(X_{\bar{K}}, \mathbb{Q}_p\right)$ from the crystalline cohomology $\mathrm{H}^n_{\mathrm{cris}}\left(Y/W, \mathcal{O}_{Y/W}\right) \otimes_W K_0$ (equipped with the extra structures), through a certain "mysterious functor".
The existence of this mysterious functor was formulated more precisely (and generally) by Fontaine.

Theorem 6.1 (Faltings) — Let X be a smooth proper variety over K which extends to a smooth model $\mathcal{X}/\mathcal{O}_K$, with special fibre Y. There is a natural functorial comparison isomorphism

$$\mathrm{C}_{\mathrm{cris}}\colon \mathrm{H}^i_{\text{ét}}\left(X_{\bar{K}}, \mathbb{Z}_p\right) \otimes_{\mathbb{Z}_p} \mathbf{B}_{\mathrm{cris}} \xrightarrow{\sim} \mathrm{H}^i_{\mathrm{cris}}(Y/W, \mathcal{O}_{Y/W}) \otimes_W \mathbf{B}_{\mathrm{cris}}.$$

Furthermore, this isomorphism extends to the level of the respective E_∞-algebras, and is compatible with the G_K action, the Frobenius operator, the filtrations (after $-\otimes_{K_0} K$), Poincaré duality, the Künneth formula, Tate twists, Chern classes of vector bundles and cycle class maps. ◇□

7 Filtered φ-modules

We would do well to wrap up the extra structures on the crystalline cohomology in a single piece of terminology. We saw in section 5.3.3 that the comparison theorem between crystalline and de Rham cohomology (Theorem 5.3) furnishes the K-vector space $\mathrm{H}^n_{\mathrm{cris}}(Y/W, \mathcal{O}_{Y/W}) \otimes_W K$ with a filtration, coming from the Hodge filtration on $\mathrm{H}^n_{\mathrm{dR}}(X/K)$. On the other hand, $\mathrm{H}^n_{\mathrm{cris}}(Y/W, \mathcal{O}_{Y/W}) \otimes_W K_0$ comes equipped with a semilinear action of a Frobenius endomorphism φ. This leads us to make the following definition:

Definition 7.1 — A **filtered φ-module** over K consists of:

- a **φ-module** (or **φ-isocrystal**) over K_0, that is:
  - a finite dimensional K_0-vector space D,
  - a σ-semilinear operator $\varphi\colon D \to D$, whose linearisation $\varphi^\sharp\colon \sigma^*(D) = D \otimes_\sigma K_0 \to D$ is an isomorphism,
- which is equipped with a finite decreasing filtration $F^\bullet$ on $D \otimes_{K_0} K$. ◇

Remark 7.2 – The terminology "φ-isocrystal" reflects that we are considering φ-crystals up to *isogeny* (see section 8B.1 for a definition of φ-crystals). Note that, for a φ-crystal M, $D = M \otimes_W K_0$ is an isocrystal; the condition that φ_M be injective corresponds to $\varphi_D^\sharp$ being an isomorphism.

The crystalline comparison theorem assures us that we can naturally recover $\mathrm{H}^i_{\text{ét}}(X_{\overline{K}}, \mathbb{Q}_p)$, *together with its Galois action*, from the filtered φ-module $\mathrm{H}^n_{\mathrm{cris}}(Y/W, \mathcal{O}_{Y/W}) \otimes_W K_0$. To see this, we define:

$$\mathrm{D}_{\mathrm{cris}}\left(\mathrm{H}^i_{\text{ét}}(X_{\overline{K}}, \mathbb{Q}_p)\right) := \left(\mathrm{H}^i_{\text{ét}}(X_{\overline{K}}, \mathbb{Q}_p) \otimes_{\mathbb{Q}_p} \mathbf{B}_{\mathrm{cris}}\right)^{\mathrm{G}_K}.$$

Note then that $\mathrm{C}_{\mathrm{cris}}$ yields an isomorphism

$$\left(\mathrm{H}^i_{\text{ét}}(X_{\overline{K}}, \mathbb{Q}_p) \otimes_{\mathbb{Q}_p} \mathbf{B}_{\mathrm{cris}}\right)^{\mathrm{G}_K} \cong \mathrm{H}^n_{\mathrm{cris}}(Y/W, \mathcal{O}_{Y/W}) \otimes_W (\mathbf{B}_{\mathrm{cris}})^{\mathrm{G}_K} = \mathrm{H}^n_{\mathrm{cris}}(Y/W, \mathcal{O}_{Y/W}) \otimes_W K_0,$$

as $(\mathbf{B}_{\mathrm{cris}})^{\mathrm{G}_K} = K_0$, as we saw in section 5.4. How do we go back? We define

$$\mathrm{V}_{\mathrm{cris}}\left(\mathrm{H}^i_{\mathrm{cris}}(Y/W, \mathcal{O}_{Y/W}) \otimes_W K_0\right) := F^0\left(\mathrm{H}^i_{\mathrm{cris}}(Y/W, \mathcal{O}_{Y/W}) \otimes_W \mathbf{B}_{\mathrm{cris}}\right)^{\varphi=1}.$$

Why do we expect this to work? Thanks to $\mathrm{C}^{-1}_{\mathrm{cris}}$ now, we have

$$F^0\left(\mathrm{H}^i_{\mathrm{cris}}(Y/W, \mathcal{O}_{Y/W}) \otimes_W \mathbf{B}_{\mathrm{cris}}\right)^{\varphi=1} \cong \mathrm{H}^i_{\text{ét}}(X_{\overline{K}}, \mathbb{Q}_p) \otimes_{\mathbb{Q}_p} F^0(\mathbf{B}_{\mathrm{cris}})^{\varphi=1} = \mathrm{H}^i_{\text{ét}}(X_{\overline{K}}, \mathbb{Q}_p),$$

this time because $F^0(\mathbf{B}_{\mathrm{cris}})^{\varphi=1} = \mathbb{Q}_p$, by Corollary 5.7.

8 Admissibility

We defined in the previous section two functors

$$\left\{\begin{array}{c}\text{Finite-dimensional continuous}\\ \mathbb{Q}_p\text{-representations of } \mathrm{G}_K\end{array}\right\} \underset{\mathrm{V}_{\mathrm{cris}}}{\overset{\mathrm{D}_{\mathrm{cris}}}{\rightleftarrows}} \{\text{Filtered } \varphi\text{-modules over } K_0\}, \tag{8.1}$$

as $\mathrm{D}_{\mathrm{cris}}(V) = (V \otimes_{\mathbb{Q}_p} \mathbf{B}_{\mathrm{cris}})^{\mathrm{G}_K}$, $\mathrm{V}_{\mathrm{cris}}(D) = F^0(D \otimes_{K_0} \mathbf{B}_{\mathrm{cris}})^{\varphi=1}$.
We can't possibly expect these to be quasi-inverse equivalences of categories: the categories on both sides are too

large!

In the next two sections (sections 8A, 8B), we will see how to repair this, so as to obtain a genuine equivalence of categories after restricting the allowable objects on both sides (Theorem 8C.1).

8A Admissibility of Galois representations

We start with the left hand side of the diagram (8.1). We cannot at all expect all finite-dimensional $\mathbb{Q}_p$-representations of G_K to be described by the data of filtered φ-modules, as somehow these only cover a "good reduction" situation. On top of that, there are much more Galois representations than can come from geometry. For instance, non-integer powers of the p-adic cyclotomic character can't ever occur in geometry.

8A.1 (F, G)-regular rings

We want to pick out, among all p-adic Galois representations of G_K, those which are "detected" by $\mathbf{B}_{\mathrm{dR}}$, or $\mathbf{B}_{\mathrm{cris}}$, ... To do this, Fontaine introduced the framework of admissible Galois representations. We start off with two conditions on rings B ensuring that the resulting functor D is well behaved; that these are relevant will come to light in the proofs of the subsequent propositions 8A.2 and 8A.3. In the following, we have in mind for instance $B = \mathbf{B}_{\mathrm{cris}}$, $G = \mathrm{G}_K$, $F = \mathbb{Q}_p$.

Definition 8A.1 (Fontaine) — Let B be a topological integral domain with continuous action of a topological group G, such that $B^G = E$ is a field. Let $F \subseteq E$ be a closed subfield. We say B is (F, G)-**regular** if the following conditions are satisfied:

- $\mathrm{Frac}(B)^G = B^G$,
- if, for some $b \in B$, the line $Fb \subseteq B$ is G-invariant, then $b \in B^\times$. ◇

Note that these conditions are automatically satisfied whenever B is a field.

Given such a regular (F, G)-ring B, we define its associated Dieudonné functor D_B, from the category of continuous finite-dimensional F-representations of G to the category of finite-dimensional E-vector spaces, by $\mathrm{D}_B \colon V \mapsto (V \otimes_F B)^G$. Attached to any $\mathrm{D}_B(V)$ is a comparison morphism

$$\alpha_{B,V} \colon \mathrm{D}_B(V) \otimes_E B = (V \otimes_F B)^G \otimes_E B \hookrightarrow V \otimes_F B \otimes_E B \stackrel{1\otimes\mu}{\to} V \otimes_F B.$$

We start with the following basic observation.

Proposition 8A.2 — Let V be an F-representation of G. If B is (F, G)-regular, then $\alpha_{B,V}$ is injective. ◇

Proof: Aiming for a contradiction, assume instead that $\alpha_{B,V}\left(\sum_{i=1}^r d_i \otimes b_i\right) = 0$, with $\{d_i \in \mathrm{D}_B(V)\}_{i=1}^r$ an E-linearly independent set, of minimal cardinality with respect to this condition. Passing to $\mathrm{D}_B(V) \otimes_E \mathrm{Frac}(B)$, using G-invariance of the d_i, and dividing by b_1, we have that

$$\alpha_{B,V}\left(d_1 \otimes 1 + \sum_{i=1}^r d_i \otimes g\left(\frac{b_i}{b_1}\right)\right) = 0$$

for all $g \in G$. Subtracting the equation with $g = 1$ from the general equation, we get

$$\alpha_{B,V}\left(\sum_{i=2}^r d_i \otimes \left(g\left(\frac{b_i}{b_1}\right) - \frac{b_i}{b_1}\right)\right) = 0$$

for all $g \in G$, which by the minimality assumption means that $g\left(\frac{b_i}{b_1}\right) = \frac{b_i}{b_1}$ for all $g \in G$, wherefore $\frac{b_i}{b_1} \in \mathrm{Frac}(B)^G$. As B is (F, G)-regular, it follows then that $\frac{b_i}{b_1} \in B^G = E$. The d_i are therefore linearly dependent, as

$$\sum_{i=1}^{r} \frac{b_i}{b_1} d_i = 0$$

is an E-linear dependence. This contradicts our initial assumption. $\square$

This incites us to think of $\alpha_{B,V}$ as telling us how much of the representation V can be seen from periods in B. This intuition is confirmed by the following elementary proposition.

Proposition 8A.3 — With the same notation as above, $\alpha_{B,V}$ is an isomorphism if and only if $\dim_E \mathrm{D}_B(V) = \dim_F V$. $\diamond$

Proof: If $\alpha_{B,V}$ is an isomorphism, comparing dimensions over $\mathrm{Frac}(B)$ gives the equality $\dim_E \mathrm{D}_B(V) = \dim_F V$. To prove the converse, we start by noting that we at least know that $\alpha_{B,V}$ is an isomorphism after tensoring up to $\mathrm{Frac}(B)$, as it is then an injective linear map between vector spaces of the same dimension, by Proposition 8A.2. We want to prove it was an isomorphism to start with; to that end, we write the matrix A of $\alpha_{B,V}$ by choosing an F-basis of V, $\{v_1, \dots, v_r\}$, and an E-basis of $\mathrm{D}_B(V)$, $\{e_1, \dots, e_r\}$. The matrix A of $\alpha_{B,V}$ is then given by $\alpha_{B,V}(e_j) = \sum_{i=1}^{r} a_{ij} v_i$, and $\det(A) \in \mathrm{Frac}(B)^\times$. We want to show that in fact $\det(A) \in B^\times$, which would show that $\alpha_{B,V}$ is an isomorphism. This is where (F, G)-regularity of B comes in: we want to show that $F \det(A) \subseteq B$ is G-invariant. Now, $\alpha_{B,V}(e_1 \wedge \dots \wedge e_r) = \det(A) v_1 \wedge \dots \wedge v_r$, and as the e_i are G-invariant, this implies that $g \in G$ acts by the scalar $\det(\rho_V(g))^{-1} \in F$ on $\det(A)$. Hence we do have that $F \det(A) \subseteq B$ is G-invariant, so that $\det(A) \in B^\times$ and $\alpha_{B,V}$ is indeed an isomorphism. $\square$

These two propositions together can be taken as justification for the definition of (F, G)-regularity. At any rate, we are led to make the following definition.

Definition 8A.4 — Let V be an F-representation of G, and B an (F, G)-regular ring. We say V is B**-admissible** if $\alpha_{B,V} \colon \mathrm{D}_B(V) \otimes_E B \hookrightarrow V \otimes_F B$ is an isomorphism. Denote the full subcategory of $\mathbf{Rep}_F(G)$ consisting of B-admissible representations as $\mathbf{Rep}_F^B(G)$. $\diamond$

The upshot of these definitions is that the category $\mathbf{Rep}_F^B(G)$ has many desired properties.

Theorem 8A.5 — The category $\mathbf{Rep}_F^B(G)$ of B-admissible F-representations of G is a sub-Tannakian category over F of the Tannakian category $\mathbf{Rep}_F(G)$, neutralised over E by the fibre functor $\mathrm{D}_B \colon \mathbf{Rep}_F^B(G) \to \mathbf{Vect}_E$. $\diamond$

8A.2 Interlude: Tannakian categories

What does this mean? Tannakian categories provide the natural framework in which to consider representations of algebraic groups. They are equipped with more structure than general categories of modules: tensor products, and duals.

We start with the following definition.

Definition 8A.6 — A category $\boldsymbol{T}$ is an F**-linear rigid abelian tensor category** if:

- $\boldsymbol{T}$ is an abelian category,
- equipped with a symmetric monoidal product $\otimes$,
- with an internal reflexive Hom-functor that is compatible with $\otimes$,
- with a given ring isomorphism $F \xrightarrow{\sim} \mathrm{End}(\mathbf{1})$. $\diamond$

How do we capture the notion that the category of vector spaces is somehow lurking in the background? We want to consider the notion of fibre functors; these are forgetful functors in disguise. We want to land in some kind of category of vector spaces, but it is particularly fruitful to allow ourselves to land somewhere else than just in F-vector spaces. With the above B-admissible representations, for instance, we want to end up in E-vector spaces, with E/F some field extension. We might as well allow ourselves to land also in a category of modules over a commutative ring, or even of sheaves over a scheme. This prompts the following definition:

Definition 8A.7 — Let $\boldsymbol{T}$ be an F-linear rigid abelian tensor category. A **fibre functor** is a $\otimes$-compatible F-linear exact functor $\omega\colon \boldsymbol{T} \to \mathbf{QCoh}(S)$ into the category of quasi-coherent sheaves on an F-scheme S.[32] ◇

We can then give a definition of Tannakian categories.

Definition 8A.8 — An F-linear rigid abelian tensor category $\boldsymbol{T}$ is said to be **Tannakian** if it admits a fibre functor $\omega\colon \boldsymbol{T} \to \mathbf{QCoh}(S)$ for some non-empty scheme S/F.[33] In addition, **T** is said to be a **neutral** Tannakian category if it admits a fibre functor $\omega\colon \boldsymbol{T} \to \mathbf{Vect}_F$.[34] ◇

8A.3 Admissible p-adic Galois representations

We summarise what Theorem 8A.5 implies about $\mathbf{Rep}_F^B(G)$ and $\mathrm{D}_B\colon \mathbf{Rep}_F^B(G) \to \mathbf{Vect}_E$.

- $\mathbf{Rep}_F^B(G)$ is a strictly full subcategory of $\mathbf{Rep}_F(G)$.
- The trivial representation $\mathbf{1} = F$ is in $\mathbf{Rep}_F^B(G)$, with $\mathrm{D}_B(F) \cong E$.
- Given $V, V_1, V_2 \in \mathbf{Rep}_F^B(G)$, we have:
  - $V_1 \oplus V_2 \in \mathbf{Rep}_F(G)$.
  - $V_1 \otimes V_2 \in \mathbf{Rep}_F(G)$, with a natural isomorphism
  $$\mathrm{D}_B(V_1 \otimes_F V_2) \cong \mathrm{D}_B(V_1) \otimes_E \mathrm{D}_B(V_2).$$
  - $\mathrm{Hom}_F(V_1, V_2) \in \mathbf{Rep}_F^B(G)$, with a natural isomorphism
  $$\mathrm{D}_B(\mathrm{Hom}_F(V_1, V_2)) \cong \mathrm{Hom}_E(\mathrm{D}_B(V_1), \mathrm{D}_B(V_2)).$$
  In particular $V^\vee \in \mathbf{Rep}_F^B(G)$, with $\mathrm{D}_B(V^\vee) \cong \mathrm{D}_B(V)^\vee$ naturally.
  - $\bigwedge^n V \in \mathbf{Rep}_F^B(G)$ and $\bigvee^n V \in \mathbf{Rep}_F^B(G)$; we naturally have $\mathrm{D}_B(\bigwedge^n V) \cong \bigwedge^n \mathrm{D}_B(V)$ and $\mathrm{D}_B(\bigvee^n V) \cong \bigvee^n \mathrm{D}_B(V)$ (ditto for other Schur functors).
- Subobjects and quotients inside $\mathbf{Rep}_F(G)$ of objects of $\mathbf{Rep}_F^B(G)$ are in $\mathbf{Rep}_F^B(G)$.
- D_B is F-linear, faithful, exact (and, as seen above, it is naturally compatible with $-\otimes-$, $\mathrm{Hom}(-,-)$, $(-)^\vee$).

To provide justification, let's consider for instance the result that $\mathbf{Rep}_F^B(G)$ is closed under taking subquotients in the ambient category $\mathbf{Rep}_F(G)$. Start with an exact sequence $0 \to V' \to V \to V'' \to 0$ in $\mathbf{Rep}_F(G)$, with $V \in \mathbf{Rep}_F^B(G)$. As D_B is left-exact (it is the composite of two left-exact functors), we obtain an exact sequence
$$0 \to \mathrm{D}_B(V') \to \mathrm{D}_B(V) \to \mathrm{D}_B(V'').$$
By Proposition 8A.2, we have that
$$\dim_E \mathrm{D}_B(V) = \dim_F V, \qquad \dim_E \mathrm{D}_B(V') \leqslant \dim_F V', \qquad \dim_E \mathrm{D}_B(V'') \leqslant \dim_F V''.$$
However, $\dim_F V = \dim_F V' + \dim_F V''$, so that the above two inequalities must in fact be equalities also. This shows B-admissibility of V' and V'', and also shows that D_B is exact (on $\mathbf{Rep}_F^B(G)$; it is not exact in general on $\mathbf{Rep}_F(G)$).

[32] The $\otimes$-compatibility condition means that ω commutes (up to isomorphism) with $\otimes$, compatibly with the commutativity, associativity and unit transformations attached to $\otimes$. Note that, from the rigidity assumption, it eventually follows that such an ω necessarily takes values in locally free sheaves of finite rank over S [Deligne. $(\boldsymbol{T}, \otimes, \omega)$, 2.7, 2.8]. It also follows (for non-empty S) that ω is faithful [Deligne. $(\boldsymbol{T}, \otimes, \omega)$, Corollaire 2.10].
[33] There is also a definition which does not require us to choose any particular ω. This involves considering the collection of all fibre functors $\mathrm{Fib}(\boldsymbol{T})$; the condition is of being able to recover $\boldsymbol{T}$ from $\mathrm{Fib}(\boldsymbol{T})$. This generalises the observation that, when given a fibre functor $\omega\colon \boldsymbol{T} \to \mathbf{Vect}_F$, we can describe $\boldsymbol{T}$ as the category of F-representations of the group $\mathrm{Aut}^{\otimes}(\omega)$ of $\otimes$-compatible automorphisms of ω. See [Deligne. $(\boldsymbol{T}, \otimes, \omega)$, Théorème 1.12].
[34] Note that, by choosing a field E such that $S(E)$ is non-empty, picking a point in $S(E)$ yields a fibre functor with values in $\mathbf{Vect}_E$ for some field extension E/F.

The other properties can be checked in a straightforward manner too.

We now want to apply this formalism to the situations of interest, e.g. taking B to be $\mathbf{B}_{\mathrm{dR}}$ or $\mathbf{B}_{\mathrm{cris}}$. As $\mathbf{B}_{\mathrm{dR}}$ is a field, it is automatically $(\mathbb{Q}_p, \mathrm{G}_K)$-regular. The situation is more difficult for $\mathbf{B}_{\mathrm{cris}}$. It is a domain (as a subring of $\mathbf{B}_{\mathrm{dR}}$). We know that $(\mathbf{B}_{\mathrm{cris}})^{\mathrm{G}_K} = \mathrm{Frac}(\mathbf{B}_{\mathrm{cris}})^{\mathrm{G}_K} = K_0$ (see section 5.4). We are left with the following.

Proposition 8A.9 (Fontaine) — Suppose $\mathbb{Q}_p \cdot b \subset \mathbf{B}_{\mathrm{cris}}$ is G_K-invariant. Then b is invertible in $\mathbf{B}_{\mathrm{cris}}$. ◇

Proof: We have that $g \cdot b = \psi(g)b$ for some character ψ of G_K. Multiplying by the appropriate power of t, we may assume that, under the natural inclusion $\mathbf{B}_{\mathrm{cris}} \subset \mathbf{B}_{\mathrm{dR}}$, b lands in $\mathbf{B}_{\mathrm{dR}}^{+} \smallsetminus \mathrm{t}\mathbf{B}_{\mathrm{dR}}^{+}$. Then $\theta(b) \in \mathbb{C}_p$ spans a 1-dimensional G_K-invariant $\mathbb{Q}_p$-subspace of $\mathbb{C}_p$. Sen theory then shows that ψ is potentially unramified. Picking then a finite extension L/K such that $\psi|_{\mathrm{G}_L}$ is unramified, we get that $\theta(b) \in (\mathbb{C}_p)^{\mathrm{I}_L} = (L^{\mathrm{nr}})^{\vee}$. This means that we can naturally consider $\theta(b)$ to be an element of $\mathbf{B}_{\mathrm{dR}}^{+}$. Write then $b' = b - \theta(b) \in \mathbf{B}_{\mathrm{dR}}^{+}$, and let i be the largest integer such that $b' \in \mathrm{t}^i\mathbf{B}_{\mathrm{dR}}^{+}$; we have $i \geqslant 1$ because $\theta(b') = \theta(b) - \theta(b) = 0$. Now $\mathbb{Q}_p \cdot b'$ is also G_K-stable, with action again given by the character ψ. Multiplying by t^{-i}, it follows that $\mathbb{Q}_p \cdot \theta(\mathrm{t}^{-i}b')$ is G_K-stable, with corresponding character $\psi\chi_p^{-i}$. Using Sen theory once again, as $\psi\chi_p^{-i}$ is most definitely not potentially unramified, necessarily $b' = 0$, so that $b = \theta(b)$ is an element of $(L^{\mathrm{nr}})^{\vee} \subset \mathbf{B}_{\mathrm{dR}}^{+}$. To show then that the inverse of b in $\mathbf{B}_{\mathrm{dR}}$ actually lies in $\mathbf{B}_{\mathrm{cris}}$, it thus suffices to show that $(L^{\mathrm{nr}})^{\vee} \cap \mathbf{B}_{\mathrm{cris}} = (K_0^{\mathrm{nr}})^{\vee}$, which follows from $(\mathbf{B}_{\mathrm{cris}})^{\mathrm{G}_L} = L_0$, $(L_0^{\mathrm{nr}})^{\vee} = (K_0^{\mathrm{nr}})^{\vee}$. □

With this knowledge in hand, we can finally constrict the full category of all p-adic Galois representations of G_K. Such a p-adic representation is said to be **de Rham** (respectively **crystalline**) if it is $\mathbf{B}_{\mathrm{dR}}$-admissible (respectively, $\mathbf{B}_{\mathrm{cris}}$-admissible). We include a table of various kinds of admissible p-adic Galois representations of G_K, together with an indication of which representations coming from geometry satisfy these admissibility conditions. For instance, the p-adic étale cohomology of a smooth projective variety over K with good reduction is crystalline, by the crystalline comparison theorem (Theorem 6.1).

period ring	admissible representations	geometric interpretation
$\overline{K}$	potentially trivial	
$(K_0^{\mathrm{nr}})^{\vee}$	unramified	
$\mathbb{C}_p$	potentially unramified	
$\mathbf{B}_{\mathrm{cris}}$	crystalline	smooth proper scheme over $\mathcal{O}_K$
$\mathbf{B}_{\mathrm{st}}$	semistable	semistable scheme over $\mathcal{O}_K$
$\mathbf{B}_{\mathrm{dR}}$	de Rham	any scheme over $\mathcal{O}_K$
$\mathbf{B}_{\mathrm{HT}}$	Hodge–Tate	

Figure I.2: Various examples of admissibility conditions for p-adic Galois representations.

Here $\mathbf{B}_{\mathrm{HT}}$ is the **ring of Hodge–Tate periods**, $\mathbf{B}_{\mathrm{HT}} = \mathbb{C}_p\left[\mathrm{t}, \mathrm{t}^{-1}\right] = \bigoplus_{i\in\mathbb{Z}} \mathbb{C}_p(i) = \mathrm{gr}_F \mathbf{B}_{\mathrm{dR}}$. It is to Hodge cohomology $\bigoplus_{p,q} \mathrm{H}^q(X, \Omega^p)$ as $\mathbf{B}_{\mathrm{dR}}$ is to de Rham cohomology. It is of course $(\mathbb{Q}_p, \mathrm{G}_K)$-regular, which can be proved using the Ax–Sen–Tate theorem and the fact that the $\mathbb{C}_p$-admissible representations are precisely the potentially unramified representations.
The period ring $\mathbf{B}_{\mathrm{st}}$ will be defined later, in section 9.10.

Note however that there is some leeway in which period rings to use. For instance, when using $\mathbf{B}_{\mathrm{cris}}$ to study p-adic representations of G_K, only the φ-stable finite dimensional K_0-subspaces of $\mathbf{B}_{\mathrm{cris}}$ are of interest. One can thus use different period rings such as $\mathbf{B}_{\max}$ or $\widetilde{\mathbf{B}}_{\mathrm{rig}}$ – which have the same finite dimensional φ-stable K_0-subspaces as $\mathbf{B}_{\mathrm{cris}}$ does – and these will serve the same function within Fontaine's framework of admissible Galois representations.

8B Admissibility of filtered φ-modules

We now consider the right hand side of diagram (8.1). The category of filtered φ-modules over K is lacking the properties we expect from a category of representations: it should be a Tannakian category, as we saw in section 8A.2, yet it is not even an abelian category, because of the presence of the filtrations! This is because morphisms are not always strictly compatible with filtrations.[35]

To fix this, we need to involve restrictions that come from the particular structure of filtered φ-modules arising in geometry. We start by investigating that.

8B.1 Frobenius and the Hodge filtration

Consider again our typical geometric setup: we have Y/κ, the special fibre of some smooth proper scheme $\mathcal{X}/\mathcal{O}_K$, $\kappa = \mathcal{O}_K/\mathfrak{m}$. Theorem 5.3 furnishes $\mathrm{H}^i_{\mathrm{cris}}(Y/W, \mathcal{O}_{Y/W})$ with a Hodge filtration of its own, the pullback of the Hodge filtration of $\mathrm{H}^i_{\mathrm{dR}}(X/K)$ under the isomorphism of the theorem; this mirrors the filtration that $\mathbf{B}_{\mathrm{cris}}$ acquires through its embedding into $\mathbf{B}_{\mathrm{dR}}$. There is however a more intrinsic definition of the Hodge filtration on crystalline cohomology, which uses the divided powers ideal sheaf $\mathcal{J}$ on $\mathrm{Cris}(Y/W)$ (or on the $\mathrm{Cris}(Y/W_n)$). We have the natural PD-filtration on the crystalline structure sheaf $\mathcal{O}_{Y/S}$, $S = \mathrm{Spec}(W_n)$, given by

$$\mathcal{O}_{Y/S} = \mathcal{J}^{[0]}_{Y/S} \supseteq \mathcal{J}^{[1]}_{Y/S} \supseteq \mathcal{J}^{[2]}_{Y/S} \supseteq \cdots .$$

This filtration yields a filtration $F^\bullet$ on crystalline cohomology:

$$F^k \mathrm{H}^i_{\mathrm{cris}}(Y/W, \mathcal{O}_{Y/W}) = \mathrm{im}\left(\mathrm{H}^i_{\mathrm{cris}}(Y/W, \mathcal{J}^{[k]}_{Y/W}) \to \mathrm{H}^i_{\mathrm{cris}}(Y/W, \mathcal{O}_{Y/W})\right).$$

Unfortunately, this filtration does not necessarily agree with the pullback, through the embedding of Theorem 5.3, of the Hodge filtration on $\mathrm{H}^i_{\mathrm{dR}}(X/K)$.[36]

An obvious question crops up: how does the Frobenius endomorphism interact with the Hodge filtration? The fundamental result is the proof of the Katz conjecture, by Mazur and Ogus. We start with the following dry lemma, which we'll attempt to tease the meaning out of.

Lemma 8B.1 — Let $M = \mathrm{H}^i_{\mathrm{cris}}(Y/W, \mathcal{O}_{Y/W})/(\text{torsion})$. Then $M/\varphi(M)$ has finite length, with

$$\lg\left({}^{\varphi(M)}/_{(p^r M \cap \varphi(M))}\right) \leqslant r h^0 + (r-1)h^1 + \cdots + h^{r-1},$$

where the $h^p = \dim_\kappa \mathrm{H}^{i-p}\left(Y/\kappa, \Omega^p_{Y/\kappa}\right)$ are the naive Hodge numbers (in degree i) of Y/κ.[37] ◇

Proof: See [Berthelot–Ogus, Lemma 8.38]. □

We can restate this in a more amenable form. The endomorphism $\varphi\colon M \to M$ is σ-semilinear, and we also know it to be injective. It thus endows M with the structure of a φ-**crystal** (over κ): M is a finite free W-module equipped with an injective σ-semilinear endomorphism.[38] We can linearise φ to $\varphi^\sharp\colon \sigma^*(M) = M \otimes_\sigma W \to M$. This bears the structure of what Mazur calls a φ-**span**: an injective W-linear homomorphism of finite free W-modules of the same rank. Given a φ-span $T\colon A \to B$, we can always diagonalise T, decomposing $B = \bigoplus_{j \geqslant 0} B_j$ such that $\mathrm{im}(T) = \bigoplus_{j\geqslant 0} p^j B_j$. This is precisely because we are now allowed to use different bases on A and B,

[35] For instance, consider two one-dimensional K-vector spaces $V = W$, with the jump in the filtration of V (respectively, W) occurring at $j = 1$ (respectively, $j = 0$). Then the identity map of vector spaces $\mathrm{id}\colon V \to W$ is a morphism of filtered vector spaces which has trivial kernel and cokernel, yet is not an isomorphism. The problem is that id is not strictly compatible with filtrations: $\mathrm{id}(F^j V) \neq \mathrm{id}(V) \cap F^j W$.

[36] This can be rectified by assuming that $\mathrm{H}^i_{\mathrm{cris}}(Y/W, \mathcal{O}_{Y/W})$ is torsion-free [Berthelot–Ogus, Lemma 8.30]. However, in general, we really need to invoke the de Rham cohomology of a generic fibre to have any hope of having enough data on the cohomology of the special fibre in order to establish a crystalline comparison theorem.

[37] As the Hodge-to-de Rham spectral sequence for Y/κ does not necessarily degenerate at E_1, these can be greater than the Hodge numbers $\mathrm{h}^{p,i-p}(Y/\kappa) := \dim_\kappa \mathrm{gr}^p_F \mathrm{H}^i_{\mathrm{dR}}(Y/\kappa)$.

[38] More generally, a φ-crystal on Y/S consists of a crystal of finite locally free $\mathcal{O}_{Y/S}$-modules $\mathcal{M}$ in $(Y/S)_{\mathrm{cris}}$ equipped with a morphism $\varphi^\sharp\colon (\mathrm{Fr}_X)^*\mathcal{M} \to \mathcal{M}$. Taking $Y = \kappa$, $S = \mathbb{Z}_{(p)}$, this corresponds to the data of a finite free $W = \mathcal{W}(\kappa)$-module M with Frobenius $\varphi^\sharp\colon \sigma^* M \to M$.

and so can use a Smith normal form; we can't do the same thing directly on the level of $\varphi\colon M \to M$. Decomposing $\mathrm{im}(T)$ in such a manner, we call the numbers $\mathrm{rk}_W B_j$ the **Newton–Hodge numbers** of the φ-span $T\colon A \to B$. Considering the φ-span $\varphi^\sharp\colon \sigma^*(M) \to M$, with Newton–Hodge numbers e_j, we have that:

$$\lg\left({}^{\varphi(M)}/_{(p^r M \cap \varphi(M))}\right) = re_0 + (r-1)e_1 + \cdots + e_{r-1}.$$

This much is clear from the Smith normal form of φ, as we have

$$\varphi(M)/_{(p^r M \cap \varphi(M))} = \bigoplus_{j \geqslant 0} p^j M_j \Big/ \bigoplus_{j \geqslant 0} p^{\max(r,j)} M_j = \bigoplus_{j=0}^{r-1} \left(p^j M_j/p^r M_j\right),$$

with $\lg\left(p^j M_j/p^r M_j\right) = (r-j)e_j$.
Putting this together with Lemma 8B.1, we obtain the inequalities

$$re_0 + (r-1)e_1 + \cdots + e_{r-1} \leqslant rh^0 + (r-1)h^1 + \cdots + h^{r-1}.$$

We can make this yet more intuitive by drawing polygons. The **Newton polygon** attached to a sequence $a_0, a_1, \ldots, a_n$ is the (convex region lying above the) graph in $\mathbb{R}^2$ consisting of consecutive segments of slope j over the interval $\left[\sum_{k=0}^{j-1} a_k, \sum_{k=0}^{j} a_k\right]$. See section 16.1 for illustration.
The inequalities above, taken together, precisely state that the Newton polygon attached to the Newton–Hodge numbers $(e_0, e_1, \ldots)$ lies on or above the Newton polygon attached to the (naive) Hodge numbers $(h^0, h^1, \ldots)$.

So far we have two polygons: the Newton polygon of the Newton–Hodge numbers of the span, and the Newton polygon of the naive Hodge numbers, which is called the naive Hodge polygon. There is a third polygon, which is directly attached to $\varphi\colon M \to M$: it is the Newton polygon of the characteristic polynomial of φ. Now, φ is not W-linear, but one can check that regardless of the choices of a matrix representative with characteristic polynomial $\sum_{j=1}^n c_j x^{n-j}$, one obtains a canonical Newton polygon, as the upper convex hull of the points $(j, \mathrm{v}_p(c_j)) \in \mathbb{R}^2$. We will see a more natural way of viewing this Newton polygon in section 8B.2.
We can easily relate the Newton polygon of a φ-crystal with that of its associated span:

Lemma 8B.2 — The Newton polygon of a φ-crystal lies above the Newton polygon attached to the Newton–Hodge numbers of its associated φ-span, with equal endpoints. ◇

Proof: See [Berthelot–Ogus, Lemma 8.40]. □

As a result, this proves the Katz conjecture, which we can state as follows:

Theorem 8B.3 (Mazur, Ogus) — Let $X/\mathcal{O}_K$ be a smooth proper scheme, with special fibre Y/κ and generic fibre X/K. The Newton polygon attached to the φ-crystal $M = \mathrm{H}^i_{\mathrm{cris}}(Y/W, \mathcal{O}_{Y/W})/(\text{torsion})$ lies on or above the Newton polygon attached to its φ-span, with equal endpoints. Moreover, these both lie over the naive Hodge polygon of $\mathrm{H}^i_{\mathrm{dR}}(Y/\kappa)$, but not necessarily with equal endpoints. ◇□

The difficulties that sometimes crop up, complicating the situation, are:

- if the crystalline cohomology $\mathrm{H}^i_{\mathrm{cris}}(Y/W, \mathcal{O}_{Y/W})$ has non-zero torsion, then the mod p dimensions can shuffle around, thanks to the universal coefficient theorem, via the exact sequence

$$0 \to \mathrm{H}^i_{\mathrm{cris}}(Y/W, \mathcal{O}_{Y/W}) \otimes_W \kappa \to \mathrm{H}^i_{\mathrm{dR}}(Y/\kappa) \to \mathrm{Tor}_1^W\left(\mathrm{H}^{i+1}_{\mathrm{cris}}(Y/W, \mathcal{O}_{Y/W}), \kappa\right) \to 0;$$

- if the Hodge-to-de Rham spectral sequence of Y/κ does not degenerate at E_1, then the Hodge numbers $\dim_\kappa \mathrm{H}^q\left(Y/\kappa, \Omega^p_{Y/\kappa}\right)$ can exceed the Hodge numbers of the Hodge filtration $\dim_\kappa \mathrm{gr}^p_F \mathrm{H}^{p+q}_{\mathrm{dR}}(Y/\kappa)$.

Once these two issues are dealt with, however, we are guaranteed the equality of the endpoints of all four polygons, arranged in decreasing order: the Newton polygon of the φ-crystal $\mathrm{H}^i_{\mathrm{cris}}(Y/W, \mathcal{O}_{Y/W})$, the Newton polygon of its associated φ-span, the naive Hodge polygon of $\mathrm{H}^i_{\mathrm{dR}}(Y/\kappa)$, the Hodge polygon of $\mathrm{H}^i_{\mathrm{dR}}(X/K)$ [Illusie. $\mathrm{H}_{\mathrm{cris}}$, Theorem 1.3.9].

8B.2 Slope filtrations

There is a natural way to contextualise these different Newton polygons, which exhibits every Newton polygon as the Newton polygon attached to a filtration. In this section, we follow [André. $F^{\geqslant\lambda}$].

8B.2.1 Slope filtrations of φ-modules

We start with the case of φ-modules. We saw in section 8B.1 that to any φ-module $\varphi \circlearrowright M$ over K_0 we can attach its Newton polygon. We first need the notion of *slope*. Assuming M is non-zero, we let $-\mu(M)$[39] be the slope between the endpoints of the Newton polygon which we attached to M in section 8B.1. Alternatively, $-\mu(M) = \frac{v_p(\det(\varphi^\sharp))}{\dim_{K_0}(M)}$. This is the **slope** of M.
One has the following property of μ:

Proposition 8B.4 — Let

$$0 \to M' \to M \to M'' \to 0$$

be a short exact sequence of (non-zero) φ-modules over K_0. Then

$$\min(\mu(M'), \mu(M'')) \leqslant \mu(M) \leqslant \max(\mu(M'), \mu(M'')),$$

with equalities being strict unless $\mu(M) = \mu(M') = \mu(M'')$. ◇

Proof: Write $-\mu(M) = {}^{\deg(M)}/_{\dim_{K_0}(M)}$, i.e. $\deg(M) = v_p(\det(\varphi_M^\sharp))$.
We have that $\dim_{K_0} M = \dim_{K_0} M' + \dim_{K_0} M''$, and $\deg(M) = \deg(M') + \deg(M'')$ by multiplicativity of the determinant in exact sequences.
This reduces the statement to the following elementary observation.
Mediant inequality. Let a', a'', b', b'' be real numbers, with b', b'' (strictly) positive. Then:

$$\min\left(\frac{a'}{b'}, \frac{a''}{b''}\right) \leqslant \frac{a'+a''}{b'+b''} \leqslant \max\left(\frac{a'}{b'}, \frac{a''}{b''}\right),$$

with equality if and only if ${}^{a'}/_{b'} = {}^{a''}/_{b''}$. □

Now, to obtain the desired filtration, we bootstrap by using **universal destabilising subobjects**. By definition, a universal destabilising subobject $U \subseteq M$ is a subobject with maximal slope in a strong sense: $\mu(U) \geqslant \mu(M')$ for every subobject $M' \subseteq M$, and if $\mu(U) = \mu(M')$ for some such M', then the inclusion of M' in M factors through U, so that M' is in fact a subobject of U.
We need to show these exist. This is a special case of [André. $F^{\geqslant\lambda}$, Lemma 3.3.2].

Lemma 8B.5 — Universal destabilising subobjects exist. ◇

Proof: We proceed by induction on $\dim_{K_0} M$. If M itself is a universal destabilising subobject of M, we are done. Otherwise, this means there is some $M' \subset M$ with $\mu(M') > \mu(M)$. Choose one such M' of maximal dimension, and let U' be its maximal destabilising subobject, by the induction hypothesis. We then have $\mu(U') > \mu(M)$, and we want to show that U' is in fact a universal destabilising subobject of M. To that end, consider an arbitrary subobject M'' of M, which we might as well assume to have no subobjects of greater slope. If $M'' \subseteq M'$, there is nothing to prove as U' is a maximal destabilising subobject of M'. So we are left to consider the case when the quotient morphism $f\colon M'' \to M/M'$ is non-zero. By maximality of the dimension of M', M/M' has no subobjects of greater slope. In particular, $\mu(\mathrm{im}(f)) \leqslant \mu(M/M')$. As $\mu(M') > \mu(M)$, we see that $\mu(M/M') < \mu(M)$ by Proposition 8B.4. Similarly, as we assumed M'' had no subobjects of greater slope, the proposition also shows that M'' has no quotients of lesser slope, so that $\mu(M'') \leqslant \mu(\mathrm{im}(f))$. Hence we have:

$$\mu(M'') \leqslant \mu(\mathrm{im}(f)) \leqslant \mu(M/M') < \mu(M) < \mu(U'),$$

so that U' is indeed a universal destabilising subobject of M. □

[39] The minus sign is included so that we eventually obtain a descending filtration (matching up with the Hodge filtration), but we could just as easily not include it and work with ascending filtrations instead.

Note that, by virtue of their universal property, universal destabilising objects are unique up to unique isomorphism.

We now attach a filtration to each φ-module M. We inductively define

$$0 = M_0 \subset M_1 \subset M_2 \subset \cdots \subset M_r = M,$$

by letting M_j be the pullback by $M \to M/M_{j-1}$ of the universal destabilising subobject of M/M_{j-1}. In particular M_1 is the universal destabilising subobject of M.
Write then $\lambda_j = \mu(M_j/M_{j-1})$; we have $\lambda_1 > \lambda_2 > \ldots > \lambda_r$ by the definition of universal destabilising subobjects together with Proposition 8B.4. We then define the **slope filtration** attached to μ, $F^{\geqslant\bullet}$, by:

$$F^{\geqslant\lambda}M = M_{\sup\{j : \lambda_j \geqslant \lambda\}}.$$

This is uniquely determined by $F^{\geqslant\lambda_j}M = M_j$, $F^{\geqslant\lambda}M = 0$ if $\lambda > \lambda_1$, and being constant on the intervals $]\lambda_{j+1}, \lambda_j]$.
Going the other way, we can recover the slope of M through the formula $\deg(M) = \sum_\lambda \lambda \cdot \dim_{K_0} \mathrm{gr}_F^\lambda M$, $\mu(M) = {}^{\deg(M)}/_{\dim_{K_0} M}$, which with this definition of deg is just the formula for the slope between the endpoints of the Newton polygon attached to a filtration.

Theorem 8B.6 (Dieudonné, Manin) — Let κ be an arbitrary field of characteristic $p > 0$, and consider the Tannakian category $\mathbf{Mod}_{K_0}(\varphi)$ of φ-modules over $K_0 = W(\kappa)[{}^1/_p]$, equipped with its (descending) slope filtration $F^{\geqslant\bullet}$ attached to μ as defined above.
If κ is perfect, then $F^{\geqslant\bullet}$ is split, i.e. there is a natural isomorphism of functors $\mathrm{gr}_F \cong \mathrm{id}$. If moreover κ is algebraically closed, then all short exact sequences split. ◇□

Remark 8B.7 – For κ algebraically closed, this implies that $\mathbf{Mod}_{K_0}(\varphi)$ is a semisimple category: every object is isomorphic to a finite direct sum of simple objects.

8B.2.2 Slope filtrations on quasi–Tannakian categories

Now, we are most interested in the case of filtered φ-modules. This category has the misfortune of not being abelian, as we remarked, because morphisms are not necessarily strictly compatible with filtrations. We weaken the definition of Tannakian categories (Definition 8A.8) by asking instead that $\boldsymbol{T}$ be only *quasi-abelian* instead of abelian. Recall that an an abelian category is an additive category, with kernels and cokernels (it is pre-abelian), such that monics and epis are strict (meaning that every monomorphism is a kernel, and every epimorphism is a cokernel). We say a pre-abelian category is **quasi-abelian** if pullbacks of strict epis are strict epis, and pushouts of strict monics are strict monics. This is weaker than being abelian, but allows us to manipulate exact sequences in much the same way. This is useful for our purposes, as for K a field, the category of filtered K-vector spaces is quasi-Tannakian, but not Tannakian.

In a quasi-abelian category, a chain $(f_i)_i$ of morphisms is said to be an exact sequence if all the morphisms f_i are strict, and consecutive morphisms satisfy the usual condition $\mathrm{im}(f_i) = \ker(f_{i+1})$. Recall that a morphism f is strict if the canonical morphism $\mathrm{coim}(f) \to \mathrm{im}(f)$ is an isomorphism; equivalently if f factors as a strict epi followed by a strict monic. As we are interested in filtrations, we also also want the notion of flags: these are simply sequences of strict monomorphisms $0 = C_0 \hookrightarrow C_1 \hookrightarrow C_2 \hookrightarrow \cdots \hookrightarrow C_r = C$ with $C_i/C_{i-1} \neq 0$.
We can then define slope filtrations in the context of quasi-Tannakian categories. Following [André. $F^{\geqslant\lambda}$, Definition 3.1.1], we start with a quasi-Tannakian category $\boldsymbol{T}$ over a field K of characteristic 0. It is endowed with a rank function rk, via $\mathrm{rk}(C) = \mathrm{tr}(\mathrm{id}\colon C \to C)$, using $\mathrm{Hom}(C, C) = C^\vee \otimes C \xrightarrow{\mathrm{tr}} \mathbf{1}$. We want to define a degree function deg, taking values in some totally ordered uniquely divisible abelian group Λ. Writing $\mathrm{sk}(\boldsymbol{T})$ for the skeleton of $\boldsymbol{T}$, inspired by the proof of Proposition 8B.4, we require:

- $\deg\colon \mathrm{sk}(\boldsymbol{T}) \to \Lambda$ is additive on short exact sequences,

- $\mu = {}^{\deg}/_{\mathrm{rk}} : \mathrm{sk}(\boldsymbol{T}) \setminus \{0\} \to \Lambda$ is such that, for any epi-monic $f : C \to D$ in $\boldsymbol{T}$, $\mu(C) \leqslant \mu(D)$.

In this new context, Proposition 8B.4 clearly continues to apply. Lemma 8B.5 also remains relevant, although the proof needs slight modification, see [André. $F^{\geqslant\lambda}$, Lemma 3.3.2]. A few notions from this proof are useful. We say an object C of $\boldsymbol{T}$ is μ-semistable if $\mu(C') \leqslant \mu(C)$ for all non-zero strict subobjects C' of C. From Proposition 8B.4, we deduce that this is equivalent to $\mu(C'') \geqslant \mu(C)$ for all non-zero strict quotients C'' of C. As a result, again using universal destabilising subobjects, we can define the slope filtration attached to the slope function $\mu : \mathrm{sk}(\boldsymbol{T}) \setminus \{0\} \to \Lambda$. This is what André calls a "decreasing functorial filtration on $\boldsymbol{T}$ by strict subobjects" which is moreover finite and left-continuous. By construction, the slope filtration

$$0 \hookrightarrow C_1 \hookrightarrow C_2 \hookrightarrow \cdots \hookrightarrow C_r = C$$

of C expresses C as a successive extension of μ-semistable objects, the C_j/C_{j-1}, of decreasing slope. This uniquely characterises the slope filtration, see [André. $F^{\geqslant\lambda}$, Proposition 4.2.2], wherein can be found another important property of the slope filtration: $\mathrm{gr}_F \circ \mathrm{gr}_F = \mathrm{gr}_F$. We can then again freely interchange the notions of slope functions and slope filtrations,[40] as we can go back by using the formula $\deg(C) = \sum_\lambda \lambda \cdot \mathrm{rk}\, \mathrm{gr}_F^\lambda C$. This gives an equivalence between slope functions and slope filtrations [André. $F^{\geqslant\lambda}$, Theorem 4.2.3].

In any such situation, we can also define Newton polygons, which are defined as expected for filtrations. Given μ and $C \in \boldsymbol{T}$, write the slope filtration of C as

$$0 = C_0 \hookrightarrow C_1 \hookrightarrow C_2 \hookrightarrow \cdots \hookrightarrow C_r = C,$$

with $\mu(C_j/C_{j-1}) = \lambda_j$. The Newton polygon of C with respect to μ is then the (convex region lying above the) piecewise linear graph in $\mathbb{R}^2$ made up of consecutive segments (starting from the origin) of increasing slopes $\lambda_r < \lambda_{r-1} < \cdots < \lambda_2 < \lambda_1$ and horizontal lengths $\mathrm{rk}(C_r/C_{r-1}), \mathrm{rk}(C_{r-1}/C_{r-2}), \ldots, \mathrm{rk}(C_2/C_1), \mathrm{rk}(C_1)$, so that the vertical length of each segment is the corresponding degree.
The subobjects C_i of C then correspond to terminal contiguous sets of segments of the Newton polygon of C (for instance, the universal destabilising subobject C_1 corresponds to the rightmost, and thus steepest, segment of the Newton polygon; C_2 corresponds to the rightmost two segments), whereas quotients of C in the dual flag (of epimorphisms) correspond to initial contiguous sets of segments.

Another interesting feature of the slope function $-\mu(M) = \frac{v_p(\det(\varphi^\sharp))}{\dim_{K_0}(M)}$ is that we defined $\deg(M)$ in terms of $\deg(\det(M))$. In general, slope functions on $\boldsymbol{T}$ that factor through the determinant $\det : \boldsymbol{T} \to \mathrm{Pic}(\boldsymbol{T})$ are called **determinantal**. In this situation we have:

Proposition 8B.8 — Let $\mu : \mathrm{sk}(\boldsymbol{T}) \setminus \{0\} \to \Lambda$ be a determinantal slope function. Let $C, D \in \boldsymbol{T}$ be non-zero objects. We have:

- $\mu(C \otimes D) = \mu(C) + \mu(D)$,
- $\mu(C^\vee) = -\mu(C)$. ◇

Proof: We have $\det(C \otimes D) = \det(C)^{\otimes \mathrm{rk}(D)} \otimes \det(D)^{\otimes \mathrm{rk}(C)}$, proving the first equality. For the second, note that $\deg(C^\vee) = \deg(\det(C^\vee)) = -\deg(\det(C)) = -\deg(C)$. □

The main result that allows us to impose an admissibility condition and recover a Tannakian category is the following [André. $F^{\geqslant\lambda}$, Proposition 8.2.6]:

Proposition 8B.9 — Let $\mu : \mathrm{sk}(\boldsymbol{T}) \setminus \{0\} \to \Lambda$ be a determinantal slope function, so that deg factors through $\delta : \mathrm{Pic}(\boldsymbol{T}) \to \Lambda$. If δ is strictly increasing (i.e. given $L' \to L$ in $\mathrm{Pic}(\boldsymbol{T})$, $\delta(L') < \delta(L)$),[41] then for all $\lambda \in \Lambda$,

[40] In full details, a slope filtration is a finite decreasing functorial filtration $F^{\geqslant\bullet}$ by strict subobjects such that $\mathrm{gr}_F \circ \mathrm{gr}_F = \mathrm{gr}_F$ and such that the function $\mu = {}^{\deg}/_{\mathrm{rk}}$ for $\deg(C) = \sum_\lambda \lambda \cdot \mathrm{rk}\, \mathrm{gr}_F^\lambda C$ is a bona-fide slope function.
[41] Note that we always have $\delta(L') \leqslant \delta(L)$, otherwise μ wouldn't be a slope function.

the full subcategory $\boldsymbol{T}(\lambda)$ of $\boldsymbol{T}$ consisting of all μ-semistable objects of slope λ (together with the zero object) is abelian, and conversely. ◇

Proof: Suppose deg is increasing, and consider a morphism $f : C \to D$ between μ-semistable objects of slope λ. It factors as

$$C \twoheadrightarrow \operatorname{coim}(f) \xrightarrow{\tilde{f}} \operatorname{im}(f) \hookrightarrow D,$$

with $\tilde{f}$ an epi-monic. We want to show that $\tilde{f}$ is an isomorphism. We now have:

$$\lambda = \mu(C) \leqslant \mu(\operatorname{coim}(f)) \leqslant \mu(\operatorname{im}(f)) \leqslant \mu(D) = \lambda,$$

so that equality holds throughout. This means that $\deg(\det(\operatorname{coim}(f))) = \deg(\det(\operatorname{im}(f)))$. Because δ is *strictly* increasing, this implies that $\det(\tilde{f})$ is a isomorphism, whence $\tilde{f}$ too.
In the other direction, start with L, L' of rank 1, and a morphism $L' \to L$, and suppose they have the same slope, so that $L^\vee \otimes L'$ has slope 0. We thus have a morphism $L^\vee \otimes L' \to \mathbf{1}$ in $\boldsymbol{T}(0)$ (as all rank 1 objects are μ-semistable). This morphism is epi-monic, and hence by the assumption that $\boldsymbol{T}(0)$ is abelian, it is an isomorphism. Hence $L \cong L'$, so that $\delta \colon \operatorname{Pic}(\boldsymbol{T}) \to \Lambda$ is indeed strictly increasing. □

We can relate the condition of being semistable of a given slope with a condition on Newton polygons too. We have the following basic observation:

Proposition 8B.10 — The following conditions are equivalent:

- D is μ-semistable of slope λ.
- D is of slope λ, and for all subobjects $C \subseteq D$, the Newton polygon of C lies on or below the line given by the graph of $y = \lambda x$ in $\mathbb{R}^2$. ◇

Proof: These are equivalent as we can recover the slope of C as the slope of the line connecting the endpoints of its Newton polygon. □

We are now just missing one crucial element: the subcategory of $\boldsymbol{T}$ consisting of μ-semistable objects of slope 0, $\boldsymbol{T}(0)$, is not necessarily Tannakian, because it is not necessarily stable under $\otimes$. We know that $\mu(C \otimes D) = \mu(C) + \mu(D)$ for any objects C, D of $\boldsymbol{T}$, however we don't know that $C \otimes D$ is μ-semistable if C and D are. If this tensor product is μ-semistable whenever C and D both are, we say that μ is **$\otimes$-multiplicative**, in which case $\boldsymbol{T}(0)$ is Tannakian.

8B.3 Admissible filtered φ-modules

In the context of filtered φ-modules over K, we have two slope functions: the slope function μ_{H} attached to the underlying filtered K-vector space, and the slope function μ_{N} attached to the underlying φ-module over K_0. It is also customary to write $t_{\mathrm{H}} = \delta_{\mathrm{H}}$, $t_{\mathrm{N}} = -\delta_{\mathrm{N}}$. In section 8B.1, we saw a natural condition to impose in this context: the Newton polygon of μ_{N} should lie on or above the Hodge polygon, which is the Newton polygon of μ_{H}. To tie this together with the previous section, we can define a third slope function $\mu = \mu_{\mathrm{H}} + \mu_{\mathrm{N}}$. This defines a determinantal slope function on the quasi-Tannakian category of filtered φ-modules over K_0.
What are the μ-semistable objects of slope 0? These are the conditions:

- $\mu(D) = 0$, i.e. $t_{\mathrm{H}}(D) = t_{\mathrm{N}}(D)$,
- for any subobject $C \subseteq D$, $\mu(C) \leqslant \mu(C)$, or equivalently $t_{\mathrm{H}}(C) \leqslant t_{\mathrm{N}}(C)$.

This justifies this following definition, originally made by Fontaine.

Definition 8B.11 — A filtered φ-module D over K is **weakly admissible** if it is μ-semistable of slope 0, i.e.

- $t_{\mathrm{H}}(D) = t_{\mathrm{N}}(D)$,

- for any subobject $C \subseteq D$, $\mathrm{t_H}(C) \leqslant \mathrm{t_N}(C)$. ◇

Note that, by Proposition 8B.10, the condition on subobjects $C \subseteq D$ is equivalent to requiring that the Newton polygon of C lies on or above its Hodge polygon (with equal endpoints for $C = D$).

The category of weakly admissible filtered φ-modules over K is then the category of μ-semistable objects of slope 0 in the quasi-Tannakian K_0-linear category of filtered φ-isocrystals over K. To finally show that this category is in fact Tannakian, as per Proposition 8B.9 and the subsequent remark, it remains to prove that a tensor product of μ-semistable objects is also μ-semistable. This is, however, a difficult result. A proof is given in [Totaro. $D \otimes D'$, Theorem 1]. This result is also an indirect consequence of a later theorem, Theorem 12.1.

8C Fontaine's theorem for crystalline representations

The upshot of these admissibility conditions is that we have suitably restricted the categories on both sides of the diagram

$$\left\{ \begin{array}{c} \text{Finite-dimensional continuous} \\ \mathbb{Q}_p\text{-representations of } \mathrm{G}_K \end{array} \right\} \underset{\mathrm{V_{cris}}}{\overset{\mathrm{D_{cris}}}{\rightleftarrows}} \left\{ \text{Filtered } \varphi\text{-modules over } K_0 \right\}. \tag{8C.1}$$

Theorem 8C.1 (Fontaine) — The functors $\mathrm{D_{cris}}$ and $\mathrm{V_{cris}}$ are quasi-inverse equivalences of Tannakian categories

$$\left\{ \begin{array}{c} \text{Finite-dimensional continuous} \\ \text{crystalline } \mathbb{Q}_p\text{-representations of } \mathrm{G}_K \end{array} \right\} \underset{\mathrm{V_{cris}}}{\overset{\mathrm{D_{cris}}}{\rightleftarrows}} \left\{ \begin{array}{c} \text{Weakly admissible filtered} \\ \varphi\text{-modules over } K_0 \end{array} \right\}. \quad \diamond\square$$

This is in fact a special case of a more general theorem (Theorem 12.1). See section 12 for more details. This theorem also allows us to explain the origin of the terminology "weakly admissible". Fontaine initially defined *admissible* filtered φ-modules as those in the essential image of the functor $\mathrm{D_{cris}}$ above. With this in mind, the constitutive assertion of the above theorem is that weakly admissible filtered φ-modules are admissible.

9 Semistability

9.1 Monodromy

With the introduction of crystalline cohomology, we were able to simmer down the generic fibre datum over K to a crystallised version over κ. Crucially, however, we needed to be in a good reduction situation. However, when the special fibre is not smooth, a complication can arise.
To get a clearer picture of the situation, we inspire ourselves from the geometric situation, where we study a space fibred over the punctured unit disc.[42] What happens is that winding around the puncture can affect the situation: there's a non-trivial monodromy. In terms of cohomology, this is the fact that the cohomologies of the fibres can form a non-trivial local system, i.e. the fundamental group of the punctured disc can act non-trivially on the cohomology of the generic fibre. This corresponds to the Gauss–Manin connection on the relative de Rham cohomology of the family. At a more elementary – and historically primordial – level, this phenomenon manifests in the behaviour of algebraic integrals, such as the following elliptic integral:

$$\int_1^\infty \frac{\mathrm{d}x}{\sqrt{x(x-1)(x-\lambda)}}.$$

The above integral is a function of $\lambda \in \mathbb{C} \smallsetminus \{0,1\}$, say f. This function is then known to satisfy the **Picard–Fuchs equation**

$$4\lambda(\lambda-1)f''(\lambda) + 4(2\lambda-1)f'(\lambda) + f(\lambda) = 0.$$

Solutions to this equation locally form a two dimensional vector space, but transporting the solutions by going in a loop around 0 or 1 transforms a solution to a different one: this is the monodromy of the Picard–Fuchs equation.[43]

At any rate, the phenomenon of non-trivial monodromy can only occur when the special fibre is singular, otherwise we end up in the situation of a fibration over a contractible base, where no surprises can happen. This means we can expect to need to adjust $\mathbf{B}_{\mathrm{cris}}^+$ for periods of varieties with bad reduction.

9.2 Complex Morse theory and the Picard–Lefschetz formula

To study the monodromy of a family over the disk with a singular fibre, we begin by understanding the topological make-up of the situation. For this, it is helpful to first consider the real situation, which is that of a family of spaces fibred over the real line. Such considerations fall within the purview of Morse theory.
Morse theory studies the topology of a compact smooth manifold X by way of a **Morse function** $f\colon X \to \mathbb{R}$: by definition, this is a smooth function whose critical points are non-degenerate, i.e. f does not vanish to second order in any direction around a critical point. In particular, the critical points of f are isolated, and for each such point p we can define the index $\mathrm{ind}_f(p)$ of f at p as the dimension of the negative eigenspace of the Hessian matrix of f at p, i.e. the number of independent directions around p along which f is decreasing.
The Morse function f then captures the homotopy type of X: the diffeomorphism type of the sublevel set $X_{\leqslant t} := f^{-1}((-\infty, t])$ does not change as long as t does not cross a critical value; around a critical value t_c, $X_{\leqslant t_c+\varepsilon}$ is homotopy equivalent to $X_{\leqslant t_c-\varepsilon}$ with an i-cell attached, for $i = \mathrm{ind}_f(p)$, with p a critical point of f with critical value t_c, assuming for simplicity that f is non-resonant, i.e. distinct critical points have distinct critical values. Moreover, non-resonant Morse functions are dense in $\mathrm{C}^\infty(X, \mathbb{R})$; this is easily seen by perturbing

[42] Recall that the underlying topological space of $\mathrm{Spec}(\mathcal{O}_K)$ is the Sierpiński space, with closed point corresponding to $\mathrm{Spec}(\kappa)$ and dense open point to $\mathrm{Spec}(K)$. In particular, the punctured unit disc's arithmetic analogue is $\mathrm{Spec}(K)$.
[43] One has the following expression for f in terms of the hypergeometric function:

$$\frac{1}{\pi}\int_1^\infty \frac{\mathrm{d}x}{\sqrt{x(x-1)(x-\lambda)}} = {}_2\mathrm{F}_1\left(1/2, 1/2, 1; \lambda\right) = \sum_{n=0}^{\infty} \binom{-1/2}{n}^2 \lambda^n,$$

as long as $|\lambda| < 1$ and $|\lambda - 1| < 1$ (in particular, this disallows λ from encircling either 0 or 1). The other period, given by the integral from $-\infty$ to 0, is then $g(\lambda) = \mathrm{i} f(-\lambda)$.

f.[44]

In the complex case, where X now is a complex manifold of dimension $n+1$, we consider a proper holomorphic function $X \to \mathbb{P}^1$, which we assume to be a non-resonant holomorphic Morse function. Now *all* non-singular fibres X_t are diffeomorphic, as we can skirt around the critical values. As for the singular fibres, an n-cell degenerates to 0 in X_t as t approaches a critical value. However, in this situation it becomes interesting to consider the behaviour of the family $\{X_t\}$ of fibres, for t varying in a small disc around a critical value t_c .

Consider for example $X = \{((z_1, z_2), \lambda) \in \mathbb{A}^2 \times \mathbb{A}^1 : z_1^2 + z_2^2 = 4\lambda\}$, with $f\colon X \to \mathbb{P}^1$ the projection onto the second component. This doesn't quite fit in the above set-up, but it is here more convenient to work in an affine situation. This is a family of Riemann surfaces parametrised by $\lambda \in \mathbb{A}^1$, except for the fibre above $\lambda = 0$, which is a union of two planes crossing in a single point. Consider then $\lambda \in \mathbb{D} = \{z \in \mathbb{C} \mid |z| < 2\}$. By the Ehresmann fibration theorem, f is a fibration above $\mathbb{D}^* = \mathbb{D} \smallsetminus \{0\}$, which in particular means that the homology of the family $\{X_\lambda : \lambda \in \mathbb{D}^*\}$ is a local system on $\mathbb{D}^*$. Picking as a basepoint $\lambda = 1$, we can understand this local system through the action of the (counterclockwise) generating loop γ around $\lambda = 0$, $[\gamma] = 1 \in \pi_1(\mathbb{D}^*, \{1\}) = \mathbb{Z}$, on the homology of X_1.

In this non-compact context, there are two homology groups to consider. These are the singular homology $\mathrm{H}_1^{\mathrm{sing}}(X_1, \mathbb{Z}) = \langle \nu \rangle$, and the Borel–Moore homology $\mathrm{H}_1^{\mathrm{BM}}(X_1, \mathbb{Z}) = \langle \eta \rangle$, as depicted in the following diagram:

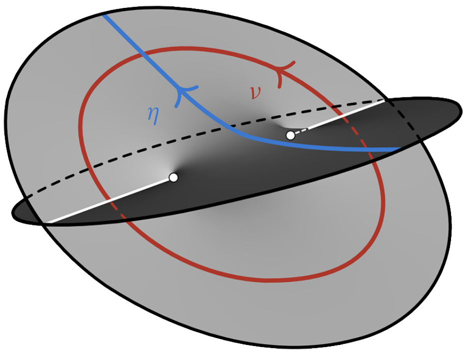


Figure I.3: A homology basis of X_1.

Note: The white lines indicate spurious self-intersections.

Then, moving λ along γ, this homology basis gets twisted as follows:

[44] In details: starting with an arbitrary smooth f, pick tubular neighbourhoods T_i over each connected component C_i of the critical set of f. Pick non-resonant Morse functions h_i with distinct critical values on each C_i, and extend them to T_i by making them constant in the normal directions. We can then perturb f to get $g = f + \varepsilon \sum \rho_i h_i$, for bump functions ρ_i which take value 1 around C_i, and then decrease in the normal direction to become 0 outside T_i. It is straightforward to see that for small enough ε, g is a non-resonant Morse function with critical points those of the h_i, with indices $\mathrm{ind}_g(p) = \mathrm{ind}_{h_i}(p) + \dim(C_i)$ (for p a critical point of h_i).

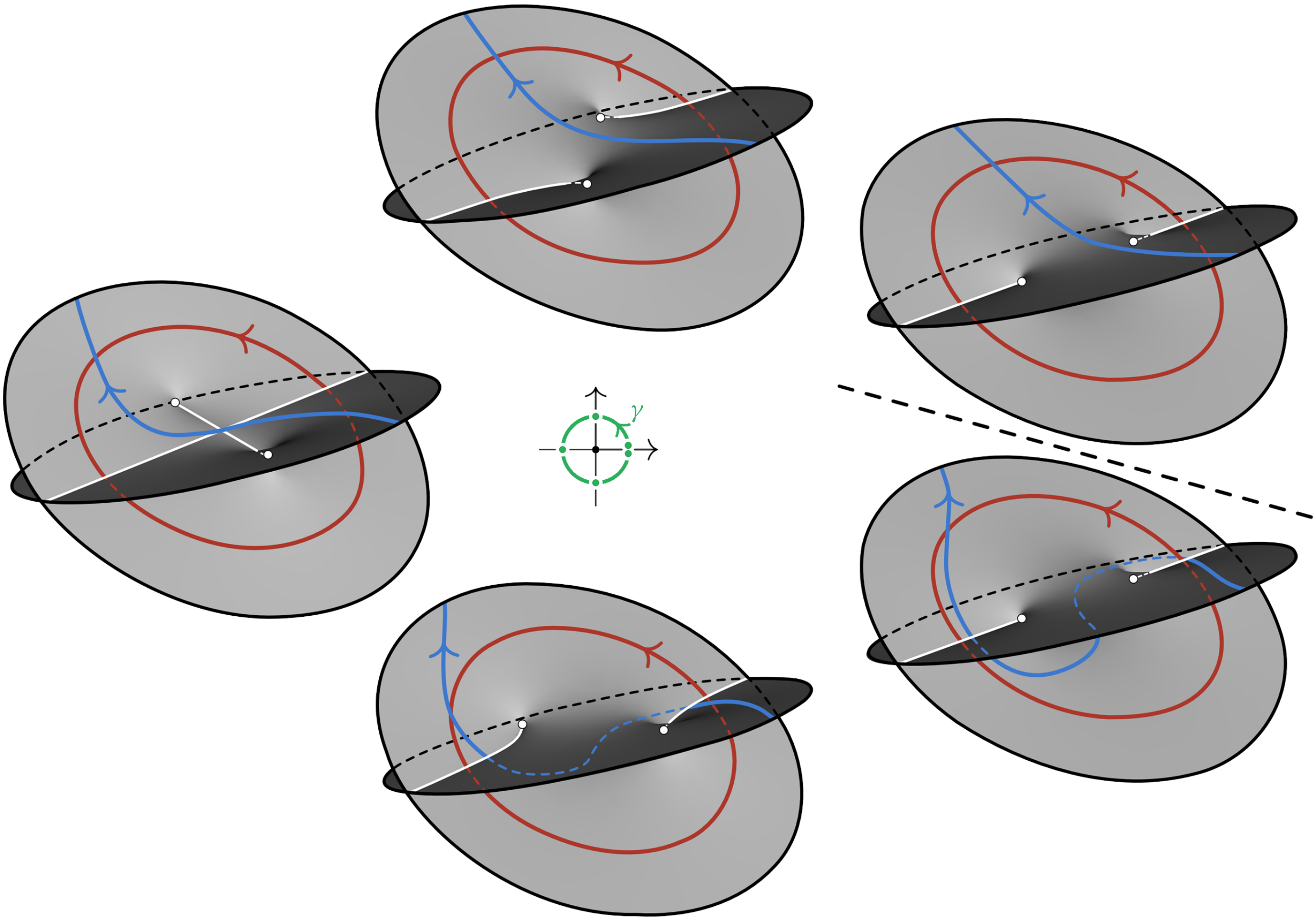


Figure I.4: Monodromy of the homology around a critical value.

We see then that $\gamma \cdot \nu = \nu$ and $\gamma \cdot \eta = \eta - \nu$; the second relation being more clearly visible if one is willing to lose the conformal structure and move apart the two sheets (flipping over the lighter-shaded sheet):

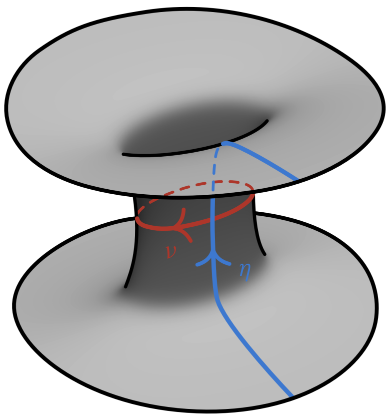


Figure I.5: A view of the topological manifold underlying X_1, with corresponding homology basis.

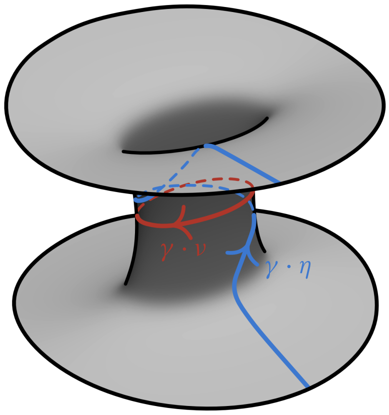


Figure I.6: The action of monodromy on the homology basis.

Topologically, the effect of γ on X_1 is then that of a Dehn twist around ν (this is the **monodromy diffeomorphism** attached to γ).[45] This is in some sense the general situation! Specifically, let $f : X \to \mathbb{D}$ be a *proper* holomorphic non-resonant Morse function with single critical value 0, with X a complex manifold of dimension 2. Write again γ for the generator of $\pi_1(\mathbb{D}^*, \{1\})$ associated with the implicit complex orientation. The monodromy diffeomorphism around 0 then corresponds to a Dehn twist around the (unique up to orientation) cycle which is contracted in the special fibre X_0 (the **vanishing cycle**). In particular, we can compute the action of $\pi_1(\mathbb{D}^*, \{1\})$ on $\mathrm{H}_1^{\mathrm{sing}}(X_1, \mathbb{Z})$:

$$\delta \mapsto \gamma \cdot \eta = \delta + \langle \nu, \delta \rangle \nu,$$

where $\delta \in \mathrm{H}_1^{\mathrm{sing}}(X_1, \mathbb{Z})$ is an arbitrary cycle, ν is a vanishing cycle for the critical value 0, and $\langle -, - \rangle$ denotes the intersection form on $\mathrm{H}_1(X_1, \mathbb{Z})$. This is the classical **Picard–Lefschetz formula**.

There is a similar formula in arbitrary dimension n, also without the assumption of properness, which computes the **variation** $\mathrm{var}(\delta) = \gamma \cdot \delta - \delta$ for $\delta \in \mathrm{H}_n^{\mathrm{BM}}(X_1, \mathbb{Z})$ using the intersection pairing with the vanishing cycle $\nu \in \mathrm{H}_n^{\mathrm{sing}}(X_1, \mathbb{Z})$ (or multiple vanishing cycles, in the case that there are multiple singular points in the special fibre X_0).

9.3 Lefschetz pencils

Now, as usual, we would like to reformulate this algebraically, without making use the notion of a holomorphic Morse function. The observation is that, around a critical point p of a Morse function f, there is always a choice of coordinates that transforms f into normal form. In the real case this is given by $f(x_1, \ldots, x_d) = -x_1^2 - \cdots - x_i^2 + x_{i+1}^2 + \cdots + x_d$, where $i = \mathrm{ind}_f(p)$, whereas in the complex case it is given by $f(z_1, \ldots, z_d) = z_1^2 + \cdots + z_d^2$. This means that the fiber X_0 over 0 resembles the (affine) quadric defined by $z_1^2 + \cdots + z_d^2 = 0$, with single singularity at $(0, \ldots, 0)$. This is the only kind of singularity we want to allow.

Let's switch to an algebraic setup now. It's useful to start by defining non-degenerate critical points algebraically. For a scheme X of finite type over a field K, and a closed point x of X, we say that x is an **ordinary quadratic singular point** if, in the case that K is algebraically closed, $\widehat{\mathcal{O}}_{X,x} \cong K[\![x_1, \ldots, x_{d+1}]\!]/(f)$, for some $f \in Q(x_1, \ldots, x_{d+1}) + \mathfrak{m}^3$, with Q a non-singular[46] quadratic form in the variables $x_1, \ldots, x_{n+1}$, and where $d = \dim_x X$. For general K, x is an ordinary quadratic singular point of X if all points above x on $X \times_K \bar{K}$ are. There's an algebraic normal form for ordinary quadratic forms. If d is even, we write $d = 2m$, and a normal form is given by $Q(x_1, \ldots, x_d) = \sum_{i=1}^m x_i x_{i+m}$; for d odd, write $d = 2m + 1$, and write a normal form $Q(x_1, \ldots, x_d) = \sum_{i=1}^m x_i x_{i+m} + x_{2m+1}^2$. Then every non-singular quadratic form over a separably closed field

[45] Note that the Dehn twist is defined independently of the orientation on ν, depending instead on the orientation already present on X_1.

[46] A quadratic form is non-singular if the associated projective quadric is non-singular. This is equivalent to non-degeneracy, except if the characteristic is 2.

can be put into normal form by a change of basis.

Consider then a smooth proper irreducible scheme X over a field K, together with a morphism $f\colon X \to \mathbb{P}^1$. We say f is a **Lefschetz pencil** if f is flat, with only finitely many singular fibres, with the singularities ordinary quadratic singular points, at most one per fibre. This is the algebraic version of the notion of a non-resonant Morse function (and in fact this notion was discovered by Solomon Lefschetz by 1924, preceding Marston Morse's 1929 paper by a few years).
To find Lefschetz pencils on X, there is again a genericity property that we can use, that will also shed light on the terminology "Lefschetz pencil". Unsurprisingly, we can use Bertini's theorem. We suppose we are given a projective embedding $\iota\colon X \hookrightarrow \mathbb{P}^N$, and consider a line $L \subset \check{\mathbb{P}}^N$ (called the screen) and associated $(d-2)$-dimensional linear subspace $\check{L} \subset \mathbb{P}^N$. In addition, we vary $P \in L$, obtaining a pencil of hyperplanes $H_P = \check{P} \supset \check{L}$. These hyperplanes all meet along the axis $\check{L}$, but are otherwise pairwise disjoint. Each point of X then lies on a unique hyperplane H_P, with the exception of the points on the base locus $B = X \cap \check{L}$. We thus get a morphism $f_L\colon X \smallsetminus B \to L \cong \mathbb{P}^1$, which we hope to prove is a Lefschetz pencil. This happens when the following conditions are satisfied:

- $\check{L}$ intersects X transversely.
- There is a Zariski-open subset $U \subseteq L$ such that the hyperplanes H_P for $P \in U$ intersect X transversely.
- For $P \in L \smallsetminus U$, H_P intersects X transversely except in a single point, an ordinary quadratic singular point of $X \cap H_P$.

In addition, we say that ι is a **Lefschetz embedding** when the subvariety of the Grassmannian $\mathrm{Gr}(1, \check{\mathbb{P}}^N)$ of the variety of lines L in $\check{\mathbb{P}}^N$ that satisfy the above three conditions is dense. Note that this subvariety is necessarily Zariski-open [SGA 7_{II}, Exposé XVII, Corollaire 3.2.1]. It is then proved [SGA 7_{II}, Exposé XVII, Théorème 2.5] that every projective embedding in characteristic 0 is a Lefschetz embedding, essentially by two applications of Bertini's theorem (one for the choice of L, another to show the second condition above). In general characteristic, it is instead proved that post-composing ι with the k-th Veronese embedding, for any $k \geqslant 2$, always yields a Lefschetz embedding.
Now, we can either work with $f_L\colon X \smallsetminus B \to \mathbb{P}^1$, or we can blow-up X at B to obtain $\widetilde{X} = \{(x, P) \in X \times L : x \in H_P\}$, so that f_L extends to $\widetilde{f}_L\colon \widetilde{X} \to \mathbb{P}^1$.

9.4 The ℓ-adic Picard–Lefschetz formula

Now that we know which objects to consider, we can try to study the monodromy algebraically. We again reduce to a local case, by use of the étale topology. Given a Lefschetz pencil $f\colon \widetilde{X} \to \mathbb{P}^1$ as above, we consider what happens around a singular point $x \in X$, so we consider the étale local ring at $f(x)$, a strictly Henselian discrete valuation ring.
We are then in the following situation (changing notation slightly): we have a flat morphism $f\colon \mathcal{X} \to S$, pure of relative dimension d, with $S = \mathrm{Spec}(\mathcal{O}_K)$ the spectrum of a Henselian discrete valuation ring of residue characteristic p, such that the generic fibre $X/\mathrm{Spec}(K)$ of f is smooth, and its special fibre $Y/\mathrm{Spec}(\kappa)$ has as only singularity $y \in Y$, an ordinary quadratic singular point. As before, we want to consider a monodromy action; in this case, instead of considering the action of a closed loop on the homology of the generic fibre, the analogous set-up is to consider the action of the étale fundamental group of S on the étale cohomology of the generic fibre.

The full situation is as depicted in the following diagram, with all squares cartesian:

$$\begin{tikzcd}
& \overline{Y} \arrow[rr, hook, "\bar{i}"] \arrow[dl] \arrow[dd] & & \overline{\mathcal{X}} \arrow[dl] \arrow[dd, "\bar{f}"] & & \overline{X} \arrow[ll, hook', "\bar{j}"'] \arrow[dl] \arrow[dd] \\
Y \arrow[rr, hook, "i"] \arrow[dd] & & \mathcal{X} \arrow[dd, "f"] & & X \arrow[ll, hook', "j"'] \arrow[dd] & \\
& \mathrm{Spec}(\overline{\kappa}) \arrow[rr, hook] \arrow[dl] & & \overline{S} \arrow[dl] & & \mathrm{Spec}(\overline{K}) \arrow[ll, hook'] \arrow[dl] \\
\mathrm{Spec}(\kappa) \arrow[rr, hook] & & S & & \mathrm{Spec}(K) \arrow[ll, hook'] &
\end{tikzcd}$$

In the notation of section 9.2, we have the dictionary $S \rightsquigarrow \mathbb{D}$, $\mathrm{Spec}(K) \rightsquigarrow \mathbb{D}^*$ (the punctured disk), $\mathrm{Spec}(\overline{\kappa}) \rightsquigarrow \{0\}$, $\mathcal{X} \rightsquigarrow X$, $Y \rightsquigarrow X_0$ (the special fibre), $X \rightsquigarrow X_1$ (the geometric generic fibre). The morphism $\mathrm{Spec}(\overline{K}) \to \mathrm{Spec}(K)$ corresponds to the universal cover of $\mathbb{D}^*$.

We are interested in looking at the étale cohomology of the geometric generic fibre. Start with a lisse ℓ-adic sheaf $\mathcal{F}$ on $\mathcal{X}$, with $\ell \neq p$ as usual. We restrict it to X, move it back to $\mathcal{X}$ by (derived) pushforward, and pull it back to Y. This defines the **nearby cycles functor**, $\mathcal{F} \mapsto \mathrm{R}\Psi_f\mathcal{F} = \bar{i}^*\mathrm{R}\bar{j}_*j^*\mathcal{F}$. This allows us to define the vanishing cycles, as the difference between $\bar{i}^*\mathcal{F}$ on Y, and the complex of sheaves $\mathrm{R}\Psi_f\mathcal{F}$. More precisely, from the adjunction morphism $\mathrm{id} \Rightarrow \mathrm{R}\bar{j}_*\bar{j}^*$ we obtain a morphism in the derived category $\bar{i}^*\mathcal{F} \to \mathrm{R}\Psi_f\mathcal{F}$, and we then define the **vanishing cycles functor** $\mathrm{R}\Phi_f$ as given by the (homotopy) cokernel of this morphism, to give a distinguished triangle

$$\bar{i}^*\mathcal{F} \longrightarrow \mathrm{R}\Psi_f\mathcal{F} \longrightarrow \mathrm{R}\Phi_f\mathcal{F}.$$

The complex of nearby cycles $\mathrm{R}\Psi_f\mathcal{F}$ is to be considered as an element in the derived category of the special fibre Y, but it pays off to also keep track of the full Galois action (that is, to remember the action of the full absolute Galois group of K, and not just of the absolute Galois group of the residue field κ). This is done by considering sheaves of the topos $Y_{\text{ét}} \times_S S_{\text{ét}}$, the (2-)fibered product of the étale toposes of Y and S, over S.[47] We are then able to extend the above definition of $\mathrm{R}\Psi$ so that it takes values in the derived category of complexes of sheaves in $Y_{\text{ét}} \times_S S_{\text{ét}}$ [SGA 7_{II}, Exposé XIII, 1.3.3].[48]

For simplicity, we restrict our choice of coefficients $\mathcal{F}$ to $\Lambda \in \{\mathbb{Z}/N\mathbb{Z}, \mathbb{Z}_\ell, \mathbb{Q}_\ell\}$ (with $p \nmid N$), so that we can dispense with j^*. We expect the cohomology of the special fibre with coefficients in $\mathrm{R}\Psi_f\Lambda$ to compute the cohomology of the generic fibre with coefficients in Λ, so that $\mathbf{H}^n_{\text{ét}}\left(\overline{Y}, \mathrm{R}\Psi_f\Lambda\right) = \mathrm{H}^n_{\text{ét}}(\overline{X}, \Lambda)$. We can deduce this from the Leray spectral sequence attached to $\bar{j}_*$, which tells us that $\mathrm{H}^n_{\text{ét}}(\overline{X}, \Lambda) = \mathbf{H}^n_{\text{ét}}(\overline{\mathcal{X}}, \mathrm{R}\bar{j}_*\Lambda)$. We just need to be able to pull this back through $\bar{i}^*$, which we can do for instance if $f\colon X \to S$ is proper, by proper base change. Thus, *when f is proper*, we have the **spectral sequence of nearby cycles**

$$\mathrm{E}_2^{p,q} = \mathrm{H}^p_{\text{ét}}\left(\overline{Y}, \mathrm{R}^q\Psi_f\Lambda\right) \Rightarrow \mathrm{H}^n_{\text{ét}}(\overline{X}, \Lambda),$$

compatible with the Galois action.

Now, using the complex of nearby cycles, we would like to be able to define a variation morphism, that captures the monodromy action of the fundamental group of $\mathrm{Spec}(K)$, i.e. the absolute Galois group of K. We are more interested in the action of the inertia subgroup I_K than of the Frobenius, as it is the inertia that corresponds to the geometric notion of going around the puncture. What we then expect is that I_K acts on the cohomology through the tame ℓ-adic fundamental character $\mathrm{t}_\ell\colon \mathrm{I}_K \to \mathbb{Z}_\ell(1)$ given by $g \cdot \zeta_{\ell^N} = \mathrm{t}_\ell(g)\zeta_{\ell_N}$; this most closely matches the geometric intuition.

The **ℓ-adic Picard–Lefschetz formula** is then the expression for the action of $g \in \mathrm{I}_K$ on $\mathrm{H}_{\text{ét}}(\overline{X}, \mathbb{Z}_\ell)$, where we

[47] The construction of this topos can be made explicit, as done in [SGA 7_{II}, Exposé XIII, 1.2], by passing to algebraic closures and describing the sheaves in terms of Galois-theoretic data: a sheaf in the topos $Y_{\text{ét}} \times_S S_{\text{ét}}$ is uniquely described by the data of a triple consisting of a sheaf $\mathcal{F}_s$ on $Y \times_S \mathrm{Spec}(\overline{\kappa})$ with continuous action of the absolute Galois group of κ (which is compatible with its action on $Y \times_S \mathrm{Spec}(\overline{\kappa})$ in the sense of [SGA 7_{II}, Exposé XIII, 1.1.1]), a sheaf $\mathcal{F}_g$ on $Y \times_S \mathrm{Spec}(\overline{\kappa})$ with continuous action of the absolute Galois group of K (again with the compatibility condition), and a Galois-equivariant morphism $\varphi\colon \mathcal{F}_s \to \mathcal{F}_g$.

[48] With the Galois-theoretic description by triples $(\mathcal{F}_s, \mathcal{F}_g, \varphi)$, we define $\Psi_f\mathcal{F}$ by giving $\mathcal{F}_s = \bar{i}^*\mathcal{F}$, $\mathcal{F}_g = \bar{i}^*\bar{j}_*\bar{j}^*\mathcal{F}$, and taking φ to be given by the adjunction morphism $\mathrm{id} \Rightarrow \bar{j}_*\bar{j}^*$, post-composed by $\bar{i}^*$. We then define $\mathrm{R}\Psi_f$ to be associated derived functor to Ψ_f.

expect that an element g with $\mathrm{t}_\ell(g) = 1$ will twist cycles by a vanishing cycle.
We want to transfer the computation of the Galois action onto the special fibre, making use of the nearby and vanishing cycle functors. Bearing in mind we are working now in a *cohomological* situation instead of a homological one, we expect the following commutative diagram [SGA 7_{II}, Exposé XIII, (2.4.6.5)] :

$$\begin{tikzcd}
\mathrm{H}^d_{\text{ét}}(\overline{X}, \mathbb{Z}_\ell) \arrow[r, "g-1\ {[49]}"] \arrow[d] & \mathrm{H}^d_{\text{ét, c}}(\overline{X}, \mathbb{Z}_\ell) \\
\mathbf{H}^d_{\text{ét, }\{y\}}(\overline{Y}, \mathrm{R}\Phi_f \mathbb{Z}_\ell) \arrow[r, "\mathrm{var}(g)_y"] & \mathbf{H}^d_{\text{ét, }\{y\}}(\overline{Y}, \mathrm{R}\Psi_f \mathbb{Z}_\ell) \arrow[u]
\end{tikzcd}$$

To define the variation morphism var, we want to use the vanishing cocycle. Dual to the topological picture, we have a generator $\breve{\nu} \in \mathbf{H}^d_{\text{ét, }\{y\}}(\overline{Y}, \mathrm{R}\Phi_f \mathbb{Z}_\ell(m)) \cong \left(\mathrm{R}^d\Phi_f\mathbb{Z}_\ell(m)\right)_y$ (where $d = 2m$ or $d = 2m+1$ depending on the parity of d). We also have a Poincaré pairing:

$$\langle -,-\rangle\colon \left(\mathrm{R}^d\Psi_f\mathbb{Z}_\ell(m)\right)_y \times \mathbf{H}^d_{\text{ét, }\{y\}}(\overline{Y}, \mathrm{R}\Psi_f\mathbb{Z}_\ell(n-m)) \longrightarrow \mathbf{H}^{2d}_{\text{ét, }\{y\}}(\overline{Y}, \mathrm{R}\Psi_f\mathbb{Z}_\ell(n)) \xrightarrow{\mathrm{tr}} \mathbb{Z}_\ell,$$

which allows us to define a class $\breve{\eta} \in \mathbf{H}^d_{\text{ét, }\{y\}}(\overline{Y}, \mathrm{R}\Psi_f\mathbb{Z}_\ell(d-m))$, the dual of $\breve{\nu}$ under the above pairing.
On the other hand, we also have a natural map $\phi\colon \mathbf{H}^d_{\text{ét, }\{y\}}(\overline{Y}, \mathrm{R}\Psi_f\mathbb{Z}_\ell) \to \mathbf{H}^d_{\text{ét, }\{y\}}(\overline{Y}, \mathrm{R}\Phi_f\mathbb{Z}_\ell)$, which behaves somewhat differently in even dimension. Referring to [SGA 7_{II}, Exposé XV, (2.2.5.3) and (2.2.5.8)], we have that:

$$\phi(\breve{\eta}) = \begin{cases} (-1)^m 2\breve{\nu} & \text{if } d \text{ is even,} \\ 0 & \text{if } d \text{ is odd.} \end{cases}$$

Because of this, the Picard–Lefschetz formula takes a different form in the case that d is even. From [SGA 7_{II}, Exposé XV, Théorème 3.4], we have:

$$\mathrm{var}(g)_y(\delta) = \begin{cases} (-1)^m \varepsilon_2(g)\langle\delta, \breve{\eta}\rangle\breve{\eta} & \text{if } d \text{ is even,} \\ (-1)^{m+1}\mathrm{t}_\ell(g)\langle\delta, \breve{\eta}\rangle\breve{\eta} & \text{if } d \text{ is odd,} \end{cases}$$

for some character $\varepsilon_2\colon \mathrm{I}_K \to \{0,1\} \cong \mathbb{Z}/2\mathbb{Z}$. Away from 2, ε_2 is the unique non-trivial mod 2 quadratic character; in characteristic 2 the situation is more subtle still [SGA 7_{II}, Exposé XV, 3.2.3].
More generally, if there are multiple quadratic ordinary singular points in the special fibre, we can add up their individual contributions. We can encode this in an operator

$$N\colon \mathrm{H}^d_{\text{ét}}(\overline{X}, \mathbb{Z}_\ell) \longrightarrow \mathrm{H}^d_{\text{ét, c}}(\overline{X}, \mathbb{Z}_\ell)(-1),$$

defined by $\mathrm{var}(g)(\delta) = N(\mathrm{t}_\ell(g))$. This operator N is simply giving us the corresponding vanishing cocycle (i.e. in the case of a single singular point, it corresponds topologically to sending $\eta \to \nu$, $\nu \to 0$). Note also that N is nilpotent; in fact $N^2 = 0$. This is because two vanishing cycles, which vanish in the same fibre, cannot be interlinked (this would contradict the assumption that the only singularities are ordinary quadratic singularities).

One particularly important situation to consider is that of the Tate elliptic curve $E_q = \mathbb{G}_{\mathrm{m},K}/q^{\mathbb{Z}}$, for some non-zero $q \in p\mathcal{O}_K$. This defines a scheme $\mathcal{X}$ over $\mathcal{O}_K$ with semistable reduction, with a single quadratic ordinary singular point on its special fibre. We can thus apply the above framework; using the above notation, we need to carefully keep track of the Tate twists in the coefficients: we have $\breve{\nu} \in (\mathrm{R}^1\Phi_f\mathbb{Z}_\ell)_y$, $\breve{\eta} \in \mathbf{H}^1_{\text{ét, }\{y\}}(\overline{Y}, \mathrm{R}\Psi_f\mathbb{Z}_\ell(1))$. The Picard–Lefschetz formula gives us that $\mathrm{var}(g)(\breve{\eta}) = 0$, $\mathrm{var}(g)(\breve{\nu}) = -\mathrm{t}_\ell(g)\breve{\eta}$. To untwist the coefficients, we have to pick an ℓ-adic orientation, i.e. a compatible system $(\zeta_{\ell^i})_i$ of primitive ℓ^i-th roots of unity. Once this is done, we can write the action of G_K on $\mathrm{H}^1(X, \mathbb{Z}_\ell)(1)$ – where the Tate twist is included for simplicity of the

[49] In the case that f is not proper, $(g-1)$ is a slight abuse of notation. To be correct, we in fact define the morphism denoted $(g-1)$ via the above commutative diagram, through the subsequent formula for $\mathrm{var}(g)_y$. Rather than proving that this formula for $\mathrm{var}(g)_y$ makes the diagram commute (which would be tautological), what is to be proved instead is that the morphism denoted $(g-1)\colon \mathrm{H}^d_{\text{ét}}(\overline{X}, \mathbb{Z}_\ell) \to \mathrm{H}^d_{\text{ét, c}}(\overline{X}, \mathbb{Z}_\ell)$ recovers the genuine action of $(g-1)$ on $\mathrm{H}^d_{\text{ét}}(\overline{X}, \mathbb{Z}_\ell)$ (respectively $\mathrm{H}^d_{\text{ét, c}}(\overline{X}, \mathbb{Z}_\ell)$) when post-composed (respectively, pre-composed) by the natural morphism $\mathrm{H}^d_{\text{ét, c}}(\overline{X}, \mathbb{Z}_\ell) \to \mathrm{H}^d_{\text{ét}}(\overline{X}, \mathbb{Z}_\ell)$.

formula (and where we also flipped the sign of $\breve{\eta}$ for convenience) – as

$$g \mapsto \begin{pmatrix} \chi_\ell(g) & \mathrm{c}_\ell(g) \\ 0 & 1 \end{pmatrix},$$

where $\mathrm{c}_\ell(g) \in \mathbb{Z}_\ell$ corresponds to the exponent of $\mathrm{t}_\ell(g)$ (so that $g \cdot \sqrt[\ell^i]{\varpi} = \mathrm{t}_\ell(g) \sqrt[\ell^i]{\varpi} = \zeta_{\ell^i}^{\mathrm{c}_\ell(g)} \sqrt[\ell^i]{\varpi}$). Note then that $\mathrm{c}_\ell(g)$ is indeed a 1-cocycle in $\mathrm{H}^1(\mathrm{G}_K, \mathbb{Z}_\ell(1))$ as it should be; that is, we have that $\mathrm{c}_\ell(gh) = \mathrm{c}_\ell(g) + \chi_\ell(g)\mathrm{c}_\ell(h)$.

9.5 Semistable schemes and mixedness

Note: In this section we work both with the complex situation (singular cohomology of complex algebraic varieties) and the ℓ-adic situation (ℓ-adic étale cohomology of schemes). In the former we can work with $\Lambda = \mathbb{Z}$, $\Lambda = \mathbb{Q}$ or $\Lambda = \mathbb{C}$, whereas in the latter we can take $\Lambda = \mathbb{Z}/\ell^N\mathbb{Z}$, $\Lambda = \mathbb{Z}_\ell$ or $\Lambda = \mathbb{Q}_\ell$, with the respective six operations formalisms. In particular, in both situations we can use the concepts of nearby and vanishing cycles, as covered in the ℓ-adic case in section 9.4.

There's another kind of singularity where we can get our hands on the monodromy. Over $\mathbb{C}$, the ordinary quadratic singularity given by $z_1^2 + z_2^2 = 0$ can be considered to be the intersection of the two lines $z_1 = \pm \mathrm{i} z_2$. So, instead of generalising to ordinary quadratic singularities in higher dimensions, an other direction for generalisation is to consider singularities that locally look like a transverse intersection of hyperplanes, i.e. that are locally given by an equation of the form $z_1 z_2 \ldots z_r = 0$.
This motivates the definition of a semistable scheme. Given $f : \mathcal{X} \to S$, with $S = \operatorname{Spec}(R)$ the spectrum of a discrete valuation ring, we say f is **strictly semistable** if the generic fibre X of $\mathcal{X}$ is smooth, and the special fibre Y is a union of smooth divisors on $\mathcal{X}$ which intersect transversally.
More specifically, around a point $y \in Y$ which lies in the intersection of r distinct divisors, the completed local ring should look like $\widehat{\mathcal{O}_{\mathcal{X},y}} \cong A[\![t_1, \ldots, t_r]\!]/(t_1 \ldots t_r - \varpi)$ for some complete local R algebra A which is formally smooth over R. This means a Zariski neighbourhood of y is smooth over $\operatorname{Spec}(R[t_1, \ldots, t_r]/(t_1 \ldots t_r - \varpi))$. We can weaken this somewhat by considering étale neighbourhoods instead, and we thus say that $\mathcal{X}$ is **semistable** if $\mathcal{X}$ is étale-locally on $\mathcal{X}$ strictly semistable, so that étale-locally around each point x of the special fibre, $\mathcal{X}$ is smooth over $\operatorname{Spec}(R[t_1, \ldots, t_r]/(t_1 \ldots t_r - \varpi))$.

How do we get a handle on the monodromy of a semistable scheme? We can reduce to the situation of a strictly semistable scheme, and then make use of the smooth divisors D_i that make up the special fibre. The divisors D_i then give us a Čech covering of Y: we have a smooth simplicial scheme $Y_\bullet$, which in degree r consists of the $(r+1)$-fold intersections of divisors in the special fibre; that is, $Y_r = \coprod_{|I|=r+1} \cap_{i \in I} D_i$, with augmentation given by $Y_{-1} = Y$. This looks like:

$$\begin{tikzcd}
\cdots & Y_3 \arrow[r, shift left=4.5] \arrow[r, shift left=1.5] \arrow[r, shift right=1.5] \arrow[r, shift right=4.5] & Y_2 \arrow[l, shift left=3] \arrow[l] \arrow[l, shift right=3] \arrow[r, shift left=3] \arrow[r] \arrow[r, shift right=3] & Y_1 \arrow[l, shift left=1.5] \arrow[l, shift right=1.5] \arrow[r, shift left=1.5] \arrow[r, shift right=1.5] & Y_0 \arrow[l] \arrow[r] & Y
\end{tikzcd}$$

We can use this Čech cover to compute the cohomology of the special fibre, with coefficients $\mathrm{R}\Psi_f\Lambda$, giving us then the cohomology of the generic fibre via the spectral sequence of nearby cycles. This is thanks to **cohomological descent**: proper hypercoverings are of cohomological descent, meaning precisely that one can use a Čech-type computation to compute cohomology [Deligne. Hodge III, §5.3].
Hopefully, if we diligently keep track of all that's going on, we should be able to get our hands on the monodromy operator too.

Now, the cohomology of the Y_i may well be complicated. What we expect to be able to describe, however, is the cycle classes attached to each D_i, and, by cup-product, also the cycle classes attached to the intersections. Writing $a_r\colon Y_r \to Y$, cohomological purity leads us to expect $\mathrm{R}^2 a_0^! \Lambda = a_0^* \Lambda(-1)$, and $\mathrm{R}^q a_0^! \Lambda = 0$ for $q \neq 2$ (in the ℓ-adic case this follows from Gabber's cohomological purity theorem).

We then want to leverage this knowledge of $\mathrm{R}^q a_{r!} \Lambda$ to compute the nearby cycles complex; the obvious tool to use is Verdier duality. This yields natural isomorphisms $\mathrm{R} i^! \Lambda = i^* \mathrm{R}\mathcal{H}\mathrm{om}(i_* \Lambda, \Lambda)$ and $\mathrm{R} j_* \Lambda = \mathrm{R}\mathcal{H}\mathrm{om}(j_! \Lambda, \Lambda)$. Using the sequence $j_! \Lambda \to \Lambda \to i_* \Lambda$ of sheaves on $\mathcal{X}$, we then put it all together to yield natural isomorphisms $i^* \mathrm{R}^q j_* \Lambda = \mathrm{R}^{q+1} i^! \Lambda = (a_{q-1})_* \mathrm{R}^{q+1} (a_{q-1})^! \Lambda = (a_{q-1})_* \Lambda(-q)$ (see [Illusie. $N \circlearrowright V/\mathbf{Q}_\ell$, Théorème 3.2][50]).

The isomorphism $i^* \mathrm{R}^q j_* \Lambda = (a_{q-1})_* \Lambda(-q)$ is thus our first ingredient. The second ingredient is the exact sequence [T. Saito. ${}_W\mathrm{E}_1^{p,q}$, Corollary 1.1.3 (2)]

$$0 \longrightarrow \mathrm{R}^q \Psi_f \Lambda \longrightarrow i^* \mathrm{R}^{q+1} j_* \Lambda(1) \longrightarrow i^* \mathrm{R}^{q+2} j_* \Lambda(2) \longrightarrow \cdots \longrightarrow i^* \mathrm{R}^{d+1} j_* \Lambda(d+1-q) \longrightarrow 0,$$

where d denotes the relative dimension of $\mathcal{X}$ over S.

Putting these two ingredients together, we obtain an explicit description of $\mathrm{R}^q \Psi_f \Lambda$, in the form of a quasi-isomorphism

$$\mathrm{R}^q \Psi_f \Lambda(q) \xrightarrow{\sim} \Big[\overset{q}{\bar{a}_{q*} \Lambda} \to \overset{q+1}{\bar{a}_{q+1*} \Lambda} \to \cdots \to \overset{d}{\bar{a}_{d*} \Lambda} \Big]$$

of complexes of sheaves on $\overline{Y}$ with Galois action. In particular, I_K acts trivially on the individual $\mathrm{R}^q \Psi_f \Lambda$. We can then use the nearby cycles spectral sequence

$$\mathrm{E}_2^{p,q} = \mathrm{H}^p \left(\overline{Y}, \mathrm{R}^q \Psi_f \Lambda \right) \Rightarrow \mathrm{H}^n (\overline{X}, \Lambda).$$

This spectral sequence is compatible with the Galois action. As I_K acts trivially on the coefficients $\mathrm{R}^q \Psi_f \Lambda$, it must also then act trivially on the associated graded of each $\mathrm{H}^n(\overline{X}, \Lambda)$ with respect to the abutment filtration. This implies that I_K acts unipotently on each $\mathrm{H}^n(\overline{X}, \Lambda)$, with index of unipotence at most $n+1$.

However we can be more precise, by describing the action of I_K directly on the level of the complex $\mathrm{R}\Psi_f \Lambda$. This will allow us to better understand how the monodromy acts on the abutment of the spectral sequence of nearby cycles.

9.5.1 Monodromy filtrations

To get us started with this problem, we begin with a brief account of monodromy filtrations.

Given a nilpotent endomorphism N of a vector space V (or more generally of an object in an abelian category), we define its **monodromy filtration** $M_\bullet$, a finite ascending filtration, as the convolution:

$$M_i V = \sum_{j-k=i} \ker(N^{j+1}) \cap \operatorname{im}(N^k).^{[51]}$$

This is the unique finite ascending filtration on V satisfying $N M_i \subseteq M_{i-2}$ such that for all i, N^i induces an isomorphism $N^i\colon \mathrm{gr}_i^M V \to \mathrm{gr}_{-i}^M V$. See [Deligne. Weil II, Proposition 1.6.1].

9.5.2 Mixed Hodge structures

Note: In this section, we are working with complex varieties, with the topological definitions for cohomology and nearby cycles.

In the complex case, for instance when we are working with $f\colon \mathcal{X} \to \mathrm{Spec}(R)$ and $R = \mathbb{C}(\!(t)\!)$, the idea is that we have two filtrations on $\mathrm{R}\Psi$: the Hodge filtration $F^\bullet$ on $\mathrm{R}\Psi_f \mathbb{C}$, and the **weight filtration** $W_\bullet$ on $\mathrm{R}\Psi_f \mathbb{Q}$, which agrees with the monodromy filtration $M_\bullet$ associated with $N\colon \mathrm{R}\Psi_f \mathbb{Q} \to \mathrm{R}\Psi_f \mathbb{Q}(-1)$. These filtrations can be defined using logarithmic differentials, as we will see in section 9.6. Accordingly, we have two spectral sequences:

[50] Note that the indexing in Illusie's article differs from ours by 1: our a_q is his a_{q+1}.

[51] This formula makes sense in an arbitrary abelian category. Indeed, given two subobjects U, V of A, we can define $U \cap V = \ker(U \to A/V)$ and $U + V = \operatorname{im}(U \oplus V \to A)$.

the **Hodge spectral sequence**

$$\mathrm{E}_1^{p,q} = \mathrm{H}^{p+q}\left(\overline{Y}, \mathrm{gr}_F^p \mathrm{R}\Psi_f \mathbb{C}\right) \Rightarrow \mathrm{H}^n_{\mathrm{sing}}(\overline{X}, \mathbb{C}).$$

and the **weight spectral sequence**

$$\mathrm{E}_1^{p,q} = \mathrm{H}^{p+q}\left(\overline{Y}, \mathrm{gr}_{-p}^W \mathrm{R}\Psi_f \mathbb{Q}\right) \Rightarrow \mathrm{H}^n_{\mathrm{sing}}(\overline{X}, \mathbb{Q}).$$

We again call the respective abutment filtrations the Hodge and weight filtrations, again denoted $F^\bullet$, $W_\bullet$.

To recapitulate then, we have the cohomology group $V = \mathrm{H}^n_{\mathrm{sing}}(\overline{X}, \mathbb{Z})$, an increasing weight filtration $W_\bullet$ defined on $V_\mathbb{Q}$, and a decreasing Hodge filtration $F^\bullet$ on $V_\mathbb{C}$. The central property of this object $(V, W_\bullet, F^\bullet)$ is that the graded pieces for the weight filtration are pure Hodge structures: $\mathrm{gr}_i^W V$, equipped with the induced Hodge filtration $F^\bullet$, is a pure Hodge structure of weight $n+i$ for each i, meaning that $F^\bullet$ is $n+i$-opposed to its complex conjugate filtration on $\mathrm{gr}_i^W V$ (which we recall means that $\mathrm{gr}_{\overline{F}}^q \mathrm{gr}_F^p \mathrm{gr}_i^W V_\mathbb{C} = 0$ for $p+q \neq n+i$). Such datum $(V, W_\bullet, F^\bullet)$, subject to the aforementioned oppositeness condition, defines a **mixed Hodge structure** of weight n [Deligne. Weil II, Définition 2.3.1]. The idea is then that the weight filtration $W_\bullet$ expresses V (or at least $V_\mathbb{Q}$) as a successive extension of pure Hodge structures.
Like in the case of pure Hodge structures, the category of mixed Hodge structures is Tannakian [Deligne. Hodge II, Théorème 2.3.5] ; the oppositeness condition is crucial as it forces strictness of the morphisms, thus ensuring the category is abelian (as, otherwise, general categories of filtered objects are not abelian). We also have a Tate mixed Hodge structures $\mathbb{Z}(1)$ (with weight filtration concentrated in degree -2). This then makes the Tannakian category of mixed Hodge structures into a **Tate triple**: it has a notion of weight, and a Tate object.

Coming back to our semistable situation, we can consider the mixed Hodge structure given by the Hodge and weight spectral sequences as a limit Hodge structure of the generic Hodge structure as $t \to 0$; the filtration $F^\bullet$ thus obtained is in general different from the usual Hodge filtration on the generic fibre (although they both yield the same filtrations on $\mathrm{gr}^W V_\mathbb{C}$ [Schmid. Variation, Corollary 6.21]). We also have additional data to work with, not just $F^\bullet$ and $W_\bullet$: we have a monodromy automorphism T.
The fundamental property of the automorphism $T \circlearrowright \mathrm{H}^n_{\mathrm{sing}}(\overline{X}, \mathbb{Z})$, in the context of a completely general flat algebraic family over the unit disc which is smooth away from 0, is that it is quasi-unipotent (so that its multiplicative Jordan decomposition $T = T_\mathrm{s} T_\mathrm{u}$ into semisimple and unipotent parts satisfies $T_\mathrm{s}^m = 1$ for some integer $m \geqslant 1$). This is Borel's monodromy theorem, which also holds in the more general setting of polarised variations of Hodge structures. In the semistable case, this result strengthens, yielding unipotence of T, so that $T = T_\mathrm{u}$. The behaviour of T is thus fully captured by its logarithm N, a nilpotent endomorphism. Steenbrink showed [Steenbrink. lim $F^\bullet$, Corollary 5.10] (with a correction by Saito [M. Saito. $\mathrm{MH}^{\mathrm{pol}}$, Remarque 4.2.5]) that in a projective situation, the monodromy filtration of the endomorphism N on $\mathrm{H}^n_{\mathrm{sing}}(\overline{X}, \mathbb{Q})$ is in fact the weight filtration (up to a renumbering); this is the **monodromy–weight theorem** for complex varieties. Steenbrink's proof proceeds by explicitly constructing, using logarithmic differentials, a bifiltered double complex $(A^{\bullet,\bullet}, W_\bullet, F^\bullet)$ whose total complex is quasi-isomorphic (as a bifiltered complex) to $(\mathrm{R}\Psi_f \mathbb{C}, W_\bullet, F^\bullet)$ [Illusie. $N \circlearrowright V/\mathbf{Q}_\ell$, Théorème 2.3.4.(a)], and studying on $A^{\bullet,\bullet}$ a certain endomorphism ν, which recovers the action of N on $\mathrm{R}\Psi_f \mathbb{C}$. See section 9.6 for (a few) more details.

The knowledge that $N\colon \mathrm{H}^n_{\mathrm{sing}}(\overline{X}, \mathbb{Q}) \to \mathrm{H}^n_{\mathrm{sing}}(\overline{X}, \mathbb{Q})(-1)$ is a morphism of mixed Hodge structures also allows us to put a bound on its index of nilpotence. The monodromy-weight theorem tells us that N^r induces an isomorphism of pure Hodge structures

$$N^r\colon \mathrm{gr}_{n+r}^W \mathrm{H}^n_{\mathrm{sing}}(\overline{X}, \mathbb{Q}) \to \mathrm{gr}_{n+r}^W \mathrm{H}^n_{\mathrm{sing}}(\overline{X}, \mathbb{Q})(-r).$$

Knowing that N is of Hodge type $(-1,-1)$, for the above isomorphism to be non-zero we certainly need a chain of non-zero subspaces:

$$F^p \mathrm{gr}^W_{n+r} \mathrm{H}^n_{\mathrm{sing}}(\overline{X}, \mathbb{C}) \xrightarrow{N} F^{p-1} \mathrm{gr}^W_{n+r-2} \mathrm{H}^n_{\mathrm{sing}}(\overline{X}, \mathbb{C}) \xrightarrow{N} \cdots \xrightarrow{N} F^{p-r} \mathrm{gr}^W_{n-r} \mathrm{H}^n_{\mathrm{sing}}(\overline{X}, \mathbb{C}).$$

Thus, for N^r to be non-zero, we at least require all the above vector spaces to be of non-zero dimension. This is precisely the requirement that the Hodge numbers $\mathrm{h}^{p,n-p}, \mathrm{h}^{p-1,n-p+1}, \ldots, \mathrm{h}^{p-r,n-p+r}$ of $\mathrm{H}^n_{\mathrm{sing}}(\overline{X}, \mathbb{Q})$ are all non-zero (this is because, as mentioned previously, the limit Hodge filtration lines up with the original Hodge filtration after taking graded pieces with respect to $W_\bullet$). It thus follows that, if m is the length of the longest chain of non-zero Hodge numbers $\mathrm{h}^{p,n-p}(\overline{X}), \mathrm{h}^{p+1,n-p-1}(\overline{X}), \ldots, \mathrm{h}^{p+m,n-m-p}(\overline{X})$, then necessarily $N^{m+1} = 0$.

9.5.3 Monodromy in ℓ-adic cohomology

The situation with ℓ-adic cohomology is somewhat different. We know to expect the Hodge filtration to be on the *algebraic de Rham cohomology*, and not on the ℓ-adic étale cohomology. The realms of coherent cohomology (where the Hodge filtration lives) and étale cohomology (where the weight filtration lives) are no longer reconciled by a comparison theorem, unless we are working over $\mathbb{C}$. So we begin with weight considerations in ℓ-adic étale cohomology.
The fundamental result in the ℓ-adic world, mirroring Borel's monodromy theorem, is Grothendieck's ℓ-adic monodromy theorem:

Theorem 9.1 (Grothendieck) — Let K be a local field of residue characteristic $p \neq \ell$, V a finite-dimensional $\mathbb{Q}_\ell$ vector space, and $\rho\colon \mathrm{G}_K \to \mathrm{GL}(V)$ an ℓ-adic Galois representation. Then $\rho|_{\mathrm{I}_K}$ is quasi-unipotent (that is, there exists a finite extension L/K such that $\rho|_{\mathrm{I}_L}$ is unipotent). ◇□

In particular, this means that there is a nilpotent operator $N\colon V \to V(-1)$ such that $\rho(g) = \exp(\mathrm{t}_\ell(g)N)$ for all $g \in \mathrm{I}_L$. This operator is in fact Galois-equivariant (explaining the need to notate it as $N\colon V \to V(-1)$ instead of simply $N\colon V \to V$). This is because of the formula $\mathrm{t}_\ell(ghg^{-1}) = \chi_\ell(g)\mathrm{t}_\ell(h)$ for $g \in \mathrm{G}_K$, $h \in \mathrm{I}_K$; it implies that $\rho(g)N\rho(g^{-1}) = \chi_\ell(g)N$, which is indeed Galois-equivariance of $N\colon V \to V(-1)$. A special case is $N\varphi = q\varphi N$, for $\varphi = \rho(\mathrm{Frob}_q^{-1})$ the image of a geometric Frobenius element.[52]
This theorem then applies to the case $V = \mathrm{H}^i_{\text{ét}}(\overline{X}, \mathbb{Q}_\ell)$ (as well as $V = \mathrm{H}^i_{\text{ét, c}}(\overline{X}, \mathbb{Q}_\ell)$). In this light it is perhaps surprising: we don't even need to know that our variety acquires semistable reduction after a finite extension; we can work purely on the level of cohomology to prove the quasi-unipotence of the monodromy.[53] We can also define a filtration $M_\bullet$ on V as the monodromy filtration of N. This is expected to have similar properties to the weight filtration of mixed Hodge structures; for instance, if $V = \mathrm{H}^i_{\text{ét}}(\overline{X}, \mathbb{Q}_\ell)$ for some smooth proper variety over K, we expect M_r/M_{r-1} to be pure of weight $i + r$ (meaning that all eigenvalues of the geometric Frobenius acting on M_r/M_{r-1} are Weil numbers of weight i).

Now, working with a semistable scheme $f\colon \mathcal{X} \to S$, we want to precise the monodromy action on $\mathrm{H}^i_{\text{ét}}(X, \mathbb{Q}_\ell)$ by making use of the spectral sequence of nearby cycles. As promised, we start by looking in depth at how the monodromy acts on $\mathrm{R}\Psi_f\Lambda$. The idea is that we again have a weight spectral sequence which mirrors the complex weight spectral sequence of Steenbrink, as constructed in [Rapoport–Zink, Satz 2.10]; see also [Illusie. $N \circlearrowright V/\mathbf{Q}_\ell$, 3.6] and [T. Saito. ${}_W\mathrm{E}_1^{p,q}$, 2.2]. One can work explicitly and construct a double complex $B^{\bullet,\bullet}$ that is an ℓ-adic analogue of Steenbrink's complex, in that its total complex is quasi-isomorphic to $\mathrm{R}\Psi_f\Lambda$. This double complex then provides a convenient way to define several structures on $\mathrm{R}\Psi_f\Lambda$.

To pull this off, recall the quasi-isomorphism of section 9.5:

$$\mathrm{R}^q\Psi_f\Lambda(q) \xrightarrow{\sim} \Big[\overset{q}{\bar{a}_{q*}\Lambda} \to \overset{q+1}{\bar{a}_{q+1*}\Lambda} \to \cdots \to \overset{d}{\bar{a}_{d*}\Lambda} \Big].$$

[52] Note that it does not matter how the arithmetic Frobenius $\mathrm{Frob}_q \in \mathrm{G}_\kappa$ is lifted to G_K, as χ_ℓ is unramified.
[53] However, we don't know that there exists a finite extension L/K that suffices for all $\ell \neq p$ simultaneously. The existence of such L/K can be proved using *alterations*, by a similar method to the proof of the implication $\mathrm{C}_{\mathrm{st}} \Rightarrow \mathrm{C}_{\mathrm{dR}}$ which will be explained at the end of section 10.

Choose now a generator T of $\mathbb{Z}_\ell(1)$. As T acts trivially on the individual $\mathrm{R}^q\Psi_f\Lambda$, we can consider that $(T-1)$ maps $\tau_{\leqslant q}\mathrm{R}\Psi_f\Lambda \to \tau_{\leqslant q-1}\mathrm{R}\Psi_f\Lambda$. Write $\widetilde{T-1}\colon \mathrm{R}^q\Psi_f\Lambda[-q] \to \mathrm{R}^{q-1}\Psi_f\Lambda[-q+1]$ for the homomorphism that $T-1$ induces after passing to graded pieces. Through the above quasi-isomorphism, $\widetilde{T-1}$ then matches up with the operation of tensoring by T [T. Saito. ${}_W\mathrm{E}_1^{p,q}$, Lemma 2.2.1 (4)]:

$$\begin{tikzcd}
0 \arrow[r] \arrow[d] & \bar{a}_{q*}\Lambda \arrow[r] \arrow[d, "\otimes T"] & \bar{a}_{q+1*}\Lambda \arrow[r] \arrow[d, "\otimes T"] & \cdots \arrow[r] & \bar{a}_{d*}\Lambda \arrow[d, "\otimes T"] \\
\bar{a}_{q-1*}\Lambda(1) \arrow[r] & \bar{a}_{q*}\Lambda(1) \arrow[r] & \bar{a}_{q+1*}\Lambda(1) \arrow[r] & \cdots \arrow[r] & \bar{a}_{d*}\Lambda(1)
\end{tikzcd}$$

The upshot of all this is that, when Λ is such that we can define logarithms (e.g. $\Lambda = \mathbb{Q}_\ell$), this description of $T-1$ also describes the action of $\log T$. This is because $\log T$ is congruent to $T-1$ modulo $(T-1)^2$, and $(T-1)^2$ sends $\tau_{\leqslant q}\mathrm{R}\Psi_f\Lambda$ to $\tau_{\leqslant q-2}\mathrm{R}\Psi_f\Lambda$, so that the morphisms induced by $T-1$ and by $\log T$ after taking graded pieces, from $\mathrm{R}^q\Psi_f\Lambda[-q]$ to $\mathrm{R}^{q-1}\Psi_f\Lambda[-q+1]$, are then equal. Being more careful with the choice of orientations, we can define $N = \log(T)\otimes T^\vee$ and $\nu = (T-1)\otimes T^\vee$ where $T^\vee \in \mathbb{Z}_\ell(-1)$ is the dual to T. The operators N and ν are then well-defined independent of choices.

From the above descriptions, we can then deduce that the kernel filtration $K_t = \ker(\nu^{t+1})$ of ν acting on $\mathrm{R}\Psi_f\Lambda$ agrees with the canonical filtration $\tau_{\leqslant\bullet}$, and we can also describe explicitly the filtration induced by the image filtration $I^i = \mathrm{im}(\nu^i)$:

$$\mathrm{gr}_{\mathrm{I}}^{t+i}\mathrm{gr}_t^{\mathrm{K}}\mathrm{R}\Psi_f\Lambda = \bigoplus_{t\geqslant \max(0,-i)} (\bar{a}_{2t+i})_*\Lambda[-2t-i](-t-i).$$

The graded pieces with respect to the monodromy filtration $M_\bullet$ of $N \circlearrowright \mathrm{R}\Psi_f\Lambda$ are thus given by

$$\mathrm{gr}_{-i}^{M}\mathrm{R}\Psi_f\Lambda = \bigoplus_{t\geqslant \max(0,-i)} \mathrm{gr}_{\mathrm{I}}^{t+i}\mathrm{gr}_t^{\mathrm{K}}\mathrm{R}\Psi_f\Lambda = \bigoplus_{t\geqslant \max(0,-i)} (\bar{a}_{2t+i})_*\Lambda[-2t-i](-t-i).$$

Assuming properness of f, thanks to the spectral sequence of nearby cycles, we can then cook up another spectral sequence, associated with the monodromy filtration $M_\bullet$ on $\mathrm{R}\Psi_f\Lambda$:

$${}_W\mathrm{E}_1^{p,q} = \mathrm{H}_{\text{ét}}^{p+q}\left(\bar{Y}, \mathrm{gr}_{-p}^{M}\mathrm{R}\Psi_f\Lambda\right) \Rightarrow \mathrm{H}_{\text{ét}}^{n}(\overline{X},\Lambda).$$

This is the **weight spectral sequence**, and it is Galois-equivariant. We just computed its first page:

$${}_W\mathrm{E}_1^{p,q} = \bigoplus_{t\geqslant \max(0,-p)} \mathrm{H}^{q-2t}(Y_{p+2t},\Lambda(-t)).$$

We can reformulate this slightly: $\mathrm{gr}^M\mathrm{R}\Psi_f\Lambda$ is the associated graded of the total complex of the double complex $B^{\bullet,\bullet}$, with all differentials 0, given by:

$$\begin{tikzcd}
\bar{a}_{d*}\Lambda & & & & \\
\bar{a}_{d-1*}\Lambda & \bar{a}_{d*}\Lambda(-1) \arrow[ul, bend right, "N"'] & & & \\
\vdots & \vdots & \ddots & & \\
\bar{a}_{1*}\Lambda & \bar{a}_{2*}\Lambda(-1) & \cdots & \bar{a}_{d*}\Lambda(-d) & \\
\bar{a}_{0*}\Lambda & \bar{a}_{1*}\Lambda(-1) \arrow[ul, bend left, "N"] & \cdots & \bar{a}_{d-1*}\Lambda(-d+1) & \bar{a}_{d*}\Lambda(-d) \arrow[ul, bend right, "N"']
\end{tikzcd}$$

We are then taking the associated graded with respect to the filtration of $\mathrm{Tot}(B)$ which in degree r is obtained by applying $\tau_{\leqslant r+q}$ to the q-th line of B (i.e. throwing away the $B^{i,q}$ for $i \geqslant r$, as the differentials are all zero); this simply corresponds to looking at anti-diagonals. We also included the monodromy endomorphism N in the

above diagram; it acts compatibly with the weight spectral sequence [Illusie. $N \circlearrowright V/\mathbf{Q}_\ell$, Théorème 2.3.4, (b)].
The abutment filtration on $\mathrm{H}^n_{\text{ét}}(\overline{X}, \Lambda)$ with respect to the weight spectral sequence is called the weight filtration, denoted $W_\bullet$. We might expect that this filtration is also the monodromy filtration of $N\colon \mathrm{H}^n_{\text{ét}}(\overline{X}, \Lambda) \to \mathrm{H}^n_{\text{ét}}(\overline{X}, \Lambda)(-1)$ (up to a renumbering); this is the (as yet unproven) **monodromy–weight conjecture**.

9.6 Log differentials

We considered ℓ-adic étale cohomology previously; we noted at the time that to understand the Hodge filtration we also need to look at de Rham cohomology. So we now switch tracks and turn to the study of differentials.
It is fruitful to consider first the absolute, non-variational problem of understanding the cohomology of possibly non-proper schemes. The method of approach is to use Nagata compactifications: embedding into a proper scheme with complement a normal crossings divisor. This set-up works over an arbitrary field of characteristic 0 (and we would need to use crystalline machinery to study the case of positive characteristic, which we will do in section 9.8). Letting K stand for an arbitrary field of characteristic 0, consider the inclusion $j\colon U \to X$ of a smooth separated scheme of finite type U over K into a smooth proper scheme X over K, with complement $D = X \smallsetminus U$ a divisor with strict normal crossings. Now, we can compute the de Rham cohomology of U as given by the hypercohomology $\mathbf{H}^n(X, j_*\Omega^\bullet_{X/K})$ of $j_*\Omega^\bullet_{U/K}$ over X; this is because j is affine, so that we have a quasi-isomorphism $j_*\Omega^\bullet_{X/K} \to \mathrm{R}j_*\Omega^\bullet_{X/K}$.

There is however a much more manageable object than $j_*\Omega^\bullet_{U/K}$ which still gives us the same de Rham cohomology; this is given by the subcomplex $\Omega^\bullet_{X/K}(\log D)$ of **logarithmic differentials**, which naturally sits in between $\Omega^\bullet_{X/K}$ and $j_*\Omega^\bullet_{U/K}$. The definition of logarithmic differentials can be done locally, using the module of logarithmic Kähler differentials. So consider the situation $X = \mathrm{Spec}(R)$, $U = \mathrm{Spec}(R[M^{-1}])$. Then the module of Kähler differentials $\Omega_{X/K}$ is generated by $\{\mathrm{d}r \mid r \in R\}$ (with usual relations), whereas $\Omega_{U/K}$ also contains differentials of the form $m^{-1}\mathrm{d}r$ for $m \in M$; in particular, poles of arbitrary degree along $D = X \smallsetminus U$ are allowed. The module of logarithmic Kähler differentials is instead generated only by $\{\mathrm{d}r \mid r \in R\}$ and $\{\mathrm{d}\log m = m^{-1}dm \mid m \in M\}$.
Geometrically, $\Omega^1_{X/K}(\log D)$ consists of those differentials α such that both α and $\mathrm{d}\alpha$ have (at most) simple poles along D. We can consider the complex of logarithmic differential $\Omega^\bullet_{X/K}(\log D)$ as generated (as a complex of quasicoherent $\mathcal{O}_X$ modules equipped with an exterior product) inside $j_*\Omega^\bullet_{U/K}$, by $\Omega^\bullet_{X/K}$ and logarithmic differential forms ${}^{\mathrm{d}f}/_f$ of sections f of $j_*\mathcal{O}_U$, rational along D [Deligne. Hodge II, Proposition 3.1.3]. Dually, we have the logarithmic tangent sheaf $\mathcal{T}_{X/K}(-\log D)$, which consists of vector fields on X which are tangent to D along D.
The main advantage of the sheaves $\Omega^i_{X/K}(\log D)$ is that they are *coherent* $\mathcal{O}_X$-modules, and in fact they are locally free in this situation, where D is a normal crossings divisor. We then have a quasi-isomorphism $\Omega^\bullet_{X/K}(\log D) \to j_*\Omega^\bullet_{U/K}$, allowing us to compute the de Rham cohomology of U using logarithmic differentials [Deligne. Hodge II, Proposition 3.1.8].[54]

Now, $\Omega^\bullet_{X/K}(\log D)$ is naturally equipped with *two* filtrations. We have a Hodge filtration $F^\bullet$ defined as the usual trivial filtration, and we also have a weight filtration $W_\bullet$:

$$W_i\Omega^p_{X/K}(\log D) := \mathrm{im}\left(\Omega^i_{X/K}(\log D) \otimes_K \Omega^{p-i}_{X/K} \longrightarrow \Omega^p_{X/K}(\log D)\right).$$

A quick topological computation shows that $\Omega^\bullet_{X/K}(\log D)$ equipped with the weight filtration $W_\bullet$ is quasi-isomorphic as a filtered complex to $\Omega^\bullet_{X/K}(\log D)$ equipped with its canonical filtration $\tau_{\leqslant\bullet}$ [Deligne. Hodge II, Proposition 3.1.8].
These two filtrations then furnish $\mathrm{H}^i_{\mathrm{dR}}(U/K) \cong \mathbf{H}^i\left(U, \Omega^\bullet_{X/K}(\log D)\right)$ with a mixed Hodge structure.

[54] Another useful quasi-isomorphic complex is the complex $\Omega^\bullet_{X/K}(*D)$ of differentials having arbitrary poles along Y. See [Deligne. Hodge II, Proposition 3.1.11].

We can now explain how to go about equipping the cohomology of any separated scheme X, of finite type over $\mathbb{C}$, with a mixed Hodge structure:

- Suppose that X is smooth. Then X embeds in a proper scheme Z over $\mathbb{C}$ [Nagata. $X \hookrightarrow \overline{X}$, §4, Theorem 2]. Moreover, one can arrange that Z is smooth, and such that the complement D of X in Z is a normal crossings divisor [Hironaka. Resolutions, §2, Main Theorem II]. Then we can equip $\mathrm{H}^i_{\mathrm{dR}}(X/\mathbb{C})$ with a mixed Hodge structure through the above isomorphism

$$\mathrm{H}^i_{\mathrm{dR}}(X/\mathbb{C}) \cong \mathbf{H}^i\left(X, \Omega^\bullet_{Z/\mathbb{C}}(\log D)\right).$$

  The crucial results are as follows [Deligne. Hodge II, Théorème 3.2.5]

  - The resulting Hodge and weight filtrations on $\mathrm{H}^i_{\mathrm{dR}}(X/\mathbb{C})$ depend only on X, and not on Z.
  - These two filtrations, together with the $\mathbb{Z}$-structure $\mathrm{H}^i_{\mathrm{sing}}(X,\mathbb{Z})$, define a mixed Hodge structure on $\mathrm{H}^i_{\mathrm{dR}}(X/\mathbb{C})$.
  - This mixed Hodge structure depends functorially on X.

- In general, dropping the assumption of smoothness, we use Hironaka's resolution of singularities to replace X by a smooth proper simplicial scheme $Z_\bullet$, normal crossing divisors $D_\bullet \hookrightarrow Z_\bullet$, such that, if $U_\bullet = Z_\bullet \smallsetminus D_\bullet$, there is an augmentation $a\colon U_\bullet \to X$ that is a proper hypercovering [Deligne. Hodge III, Proposition 8.3.2]. The mixed Hodge structures on $j\colon U_n \hookrightarrow Z_n$ then fit compatibly to yield a cohomological mixed Hodge complex[55] [Deligne. Hodge III, 8.1.12]

$$\left(\mathrm{R}j_*\mathbb{Z}, (\mathrm{R}j_*\mathbb{Q}, \tau_\leqslant), (\Omega^\bullet_{Z_\bullet}(\log D_\bullet), W, F)\right).$$

  This data then yields a functorial mixed Hodge structure on $\mathrm{H}^n_{\mathrm{dR}}(U_\bullet/\mathbb{C})$ [Deligne. Hodge III, Proposition 8.1.20]. Finally, we can use cohomological descent along the proper hypercovering $a\colon U_\bullet \to X$ to obtain a mixed Hodge structure on $\mathrm{H}^n_{\mathrm{dR}}(X/\mathbb{C})$. This is independent of all choices, and is functorial in X [Deligne. Hodge III, Proposition 8.2.2]. More generally, this method applies to any separated simplicial scheme of finite type over $\mathbb{C}$.

In the case of a semistable family of varieties $f\colon X \to S$ over the unit disc $\Delta = S$ in $\mathbb{C}$, we can also use the complex of log differentials to describe the monodromy of the Gauss–Manin connection ∇. This will generalise the description of the Gauss–Manin connection of section 1.2.1, which assumed smoothness of f.
We start by generalising to the logarithmic case the exact sequence whose connecting homomorphism defined the Gauss–Manin connection, as we saw in the proof of Theorem 1.2. Writing $U = X \smallsetminus D$, $f|_U$ is smooth, so that we have an exact sequence

$$0 \longrightarrow \mathrm{K}^1\Omega^\bullet_{U/\mathbb{C}}/\mathrm{K}^2\Omega^\bullet_{U/\mathbb{C}} \longrightarrow \mathrm{K}^0\Omega^\bullet_{U/\mathbb{C}}/\mathrm{K}^2\Omega^\bullet_{U/\mathbb{C}} \longrightarrow \mathrm{K}^0\Omega^\bullet_{U/\mathbb{C}} \longrightarrow 0.$$

Now we are working with the specific base of the punctured unit disc Δ^*, which is of dimension 1, so that by the definition of the Koszul filtration $\mathrm{K}^\bullet$, the above sequence simplifies to

$$0 \longrightarrow (f|_U)^*\left(\Omega^1_{\Delta^*/\mathbb{C}}\right) \otimes_{\mathcal{O}_U} \Omega^\bullet_{U/\Delta^*}[-1] \longrightarrow \Omega^\bullet_{U/\mathbb{C}} \longrightarrow \Omega^\bullet_{U/\Delta^*} \longrightarrow 0.$$

In the logarithmic setup, we can extend this sequence over the origin [Steenbrink. lim $F^\bullet$, 2.19], to obtain:

$$0 \longrightarrow f^*\left(\Omega^1_{S/\mathbb{C}}(\log\{0\})\right) \otimes_{\mathcal{O}_X} \Omega^\bullet_{X/S}(\log D)[-1] \longrightarrow \Omega^\bullet_{X/\mathbb{C}}(\log D) \longrightarrow \Omega^\bullet_{X/S}(\log D) \longrightarrow 0.$$

We can then use this sequence to extend the Gauss–Manin connection over the whole of S, given by the connecting homomorphism associated with the above short exact sequence after applying $\mathbf{R}^q f_*$:

$$\nabla\colon \mathbf{R}^q f_*\Omega^\bullet_{X/S}(\log D) \longrightarrow \left(\Omega^1_{S/\mathbb{C}}(\log\{0\})\right) \otimes_{\mathcal{O}_S} \mathbf{R}^q f_*\Omega^\bullet_{X/S}(\log D).$$

[55] A cohomological mixed Hodge complex is the derived version of a mixed Hodge structure, i.e. corresponds to data that precedes the application of $\mathrm{R}\Gamma$. Such an object consists of the expected data at the derived category level, satisfying the usual oppositeness condition [Deligne. Hodge III, 8.1.6].

This is an extension of the Gauss–Manin connection on $\Omega^\bullet_{U/\Delta^*}$ to a connection with log poles on $\Omega^\bullet_{X/S}(\log D)$. Now that we have extended ∇ over to D, we can explicitly study its monodromy around $z = 0$. To detect this, we take residues. We have a residue homomorphism $\mathrm{res}_0 \colon \Omega^1_{S/\mathbb{C}}(\log\{0\}) \to \mathbb{C}$ given by $\mathrm{res}_0\left(f \cdot \frac{\mathrm{d}z}{z}\right) = f(0)$. As a result, we obtain a short exact sequence

$$0 \longrightarrow \Omega^\bullet_{X/S}(\log D)[-1] \xrightarrow{\alpha\mapsto \mathrm{dlog}(z)\wedge\alpha} \Omega^\bullet_{X/\mathbb{C}}(\log D) \longrightarrow \Omega^\bullet_{X/S}(\log D) \longrightarrow 0, \tag{9.6.1}$$

allowing us then to define the residue at 0 of the Gauss–Manin connection as the associated connecting homomorphism

$$\mathrm{res}_0\nabla \colon \mathbf{R}^q f_*\Omega^\bullet_{X/S}(\log D)_{\{0\}} \longrightarrow \mathbf{R}^q f_*\Omega^\bullet_{X/S}(\log D)_{\{0\}}.$$

The fundamental consideration in the semistable situation over $\mathbb{C}$ is that we have at our disposition both $\mathrm{R}\Psi_f\mathbb{C}$ and $\Omega^\bullet_{X/S}(\log D)$, which are serving the similar purpose of describing the limit mixed Hodge structure of the family $f \colon X \to S$. The link was established by Steenbrink using his aforementioned double complex $A^{\bullet,\bullet}$, which we can finally define. The components are given by

$$A^{p,q} = \Omega^{p+q+1}_{X/K}(\log D)/W_q\Omega^{p+q+1}_{X/K}(\log D),$$

with differentials $A^{p,q} \to A^{p+1,q}$ given by $\alpha \mapsto (-1)^{q+1}\mathrm{d}\alpha$, and $A^{p,q} \to A^{p,q+1}$ given by $\alpha \mapsto (-1)^p f^*(\mathrm{dlog}(z)) \wedge \alpha$. The interest of this double complex is, as mentioned, that we can define the Hodge and weight filtrations on it in a straightforward manner. The Hodge filtration is the trivial filtration on the first index $F^i = \sigma^{\mathrm{hor}}_{\geqslant i}$, whereas the weight filtration is given by

$$W_i A^{p,q} = W_{2q+i+1}\Omega^{p+q+1}_{X/K}(\log D)/W_q\Omega^{p+q+1}_{X/K}(\log D).$$

As mentioned previously, the total complex of this bifiltered double complex $(A^{\bullet,\bullet}, F^\bullet, W_\bullet)$ is quasi-isomorphic to $(\mathrm{R}\Psi_f\mathbb{C}, F^\bullet, W_\bullet)$. Moreover, it makes explicit the action of T: the action of its logarithm N on $\mathrm{R}\Psi_f\mathbb{C}$ lines up with the action of the endomorphism $\nu \colon A \to A$ of bidegree $(-1, 1)$ induced in degree (p, q) by $(-1)^{p+q+1}$ times the natural morphism

$$\Omega^{p+q+1}_{X/K}(\log D)/W_q\Omega^{p+q+1}_{X/K}(\log D) \to \Omega^{p+q+1}_{X/K}(\log D)/W_{q+1}\Omega^{p+q+1}_{X/K}(\log D).$$

We refer to [Illusie. $N \circlearrowright V/\mathbf{Q}_\ell$, Théorème 2.3.4] for proofs and further justifications.
The upshot is then that we can deduce the existence of a quasi-isomorphism

$$\Omega^\bullet_{X/S}(\log D) \otimes_{\mathcal{O}_X} \mathcal{O}_D \xrightarrow{\sim} \mathrm{R}\Psi_f\mathbb{C}$$

in the derived category of sheaves on D [Steenbrink. lim $F^\bullet$, Lemma 4.12]. This correspondence between $\mathrm{R}\Psi_f\mathbb{C}$ and $\Omega^\bullet_{X/S}(\log D)$ then links up the Gauss–Manin connection ∇ with the logarithm of monodromy endomorphism N:

$$N = 2\pi\mathrm{i}\,\mathrm{res}_0\nabla.$$

9.7 Log geometry

In the previous section, we looked at the situation of a compactification $j \colon U \hookrightarrow X$ over a fixed base: we were compactifying $U/\mathrm{Spec}(K)$ by another scheme over $\mathrm{Spec}(K)$. However, we are more interested in spreading out some $X/\mathrm{Spec}(K)$ to $\mathcal{X}/S$ (with $S = \mathrm{Spec}(\mathcal{O}_K)$), and studying the cohomology of X through the cohomology of the special fibre Y of $\mathcal{X}$.
In section 9.5, we saw that the cohomology of the generic fibre of a semistable scheme gains a monodromy operator. In the good reduction situation, we were able to capture all the arithmetic data with only the special fibre. We can't hope for as much in this instance, but we would still like to be able to define some sort of semistable crystalline cohomology, which would have both a Frobenius operator and a monodromy operator. To do this, we introduce the notion of **log schemes**, which is a framework that mimics the consideration of logarithmic differentials of section 9.6.

The idea of Fontaine and Illusie is to axiomatise the structure we obtain from a compactification $j : U \hookrightarrow X$. We consider the subsheaf $\mathcal{M} \subseteq \mathcal{O}_X$ consisting of sections of $\mathcal{O}_X$ which are invertible on U; more precisely we define $\mathcal{M} = \mathcal{O}_X \cap j_* \mathcal{O}_U^\times$. This is a subsheaf of (commutative) monoids of the (commutative) monoid $\mathcal{O}_X$ under multiplication; unfortunately we have to throw away the additive structure as the sum of invertible sections might well not be invertible.

Definition 9.2 — Let X be a scheme. A **log structure** on X is a sheaf $\mathcal{M}$ of commutative monoids on the étale site $\text{Ét}(X)$ of X, together with a morphism of sheaves of monoids $\alpha : \mathcal{M} \to \mathcal{O}_X$ which isomorphically maps $\alpha^{-1}(\mathcal{O}_X^\times)$ onto $\mathcal{O}_X^\times$.
A **log scheme** is a scheme equipped with a log structure. ◇

The condition $\alpha^{-1}(\mathcal{O}_X^\times) \cong \mathcal{O}_X^\times$ enables us to consider $\mathcal{O}_X^\times$ as a subsheaf of $\mathcal{M}$. The reason to consider the étale site of X, and not just its Zariski site, is to avoid any problems with normal crossing divisors that are not strict; allowing étale covers leads to a simple local theory.

Next we define the category of log schemes. Consider a morphism of schemes $f : X \to Y$, and log structures $\mathcal{M}, \mathcal{N}$ on X, Y. We get morphisms $f^{-1}\mathcal{N} \to f^{-1}\mathcal{O}_Y \to \mathcal{O}_X$, and thus a composite morphism $\beta : f^{-1}\mathcal{N} \to \mathcal{O}_X$. However $f^{-1}\mathcal{N}$ does not necessarily define a log structure; we canonically correct this by instead taking the pushout of $f^{-1}\mathcal{N}$ and $\mathcal{O}_X^\times$ along $\beta^{-1}\mathcal{O}_X^\times$; this is the **associated log structure** $\left(f^{-1}\mathcal{N}\right)^{\mathrm{a}}$, which we also write $f^*\mathcal{N}$.[56] This allows us to define morphisms of log schemes as pairs: a morphism $f : X \to Y$ of schemes, and a morphism $f^\times : f^*\mathcal{N} \to \mathcal{M}$ of log structures on X.

The next step is to define sheaves of differentials in this setting. Changing notation somewhat, let $f : X \to Y$ be a morphism of log schemes, and let $\underline{X}$, $\underline{Y}$ be the underlying schemes of X, Y, with respective log structures $\mathcal{M}, \mathcal{N}$. We expect the sheaf of relative log differentials of X/Y, $\Omega^1_{X/Y}$, to be generated by the underlying differentials $\Omega^1_{\underline{X}/\underline{Y}}$ together with the logarithmic derivatives of elements in $\mathcal{M}$. That is, we expect universal morphisms $\mathrm{d} : \mathcal{O}_{\underline{X}} \to \Omega^1_{X/Y}$, $\mathrm{dlog} : \mathcal{M} \to \Omega^1_{X/Y}$. The morphism d should be a derivation, whereas dlog should be a **log derivation** of X over Y with respect to d. This means the following identities should be satisfied:

$$\begin{aligned} \mathrm{dlog}(ab) &= \mathrm{dlog}(a) + \mathrm{dlog}(b), \\ \alpha(a)\mathrm{dlog}(a) &= \mathrm{d}(\alpha(a)), \\ \mathrm{dlog}(a) &= 0 \quad \text{for } a \in f^*\mathcal{N}. \end{aligned}$$

These are the expected identities if we are to understand dlog as given by a symbolic formula $\mathrm{dlog}(x) = {}^{\mathrm{d}x}/_{x}$. This allows us to define the log tangent sheaf $\mathcal{T}_{X/Y}$ as the sheaf of log derivations[57] $\mathcal{T}_{X/Y} = \mathfrak{Der}_{X/Y}(\mathcal{O}_{\underline{X}})$. As in the classical case, the functor $\mathcal{L} \mapsto \mathfrak{Der}_{X/Y}(\mathcal{O}_{\underline{X}})$ that assigns to a coherent $\mathcal{O}_{\underline{X}}$-module its associated sheaf of log derivations is represented by a universal object, written $\Omega^1_{X/Y}$. This means that we naturally have

$$\mathcal{H}\mathrm{om}\left(\Omega^1_{X/Y}, \mathcal{L}\right) \cong \mathfrak{Der}_{X/Y}(\mathcal{L}).$$

In fact, we can also describe $\Omega^1_{X/Y}$ explicitly [Ogus. Log notes, Proposition 1.1.6, Remark 1.1.8] :

$$\Omega^1_{X/Y} = \left(\Omega^1_{\underline{X}/\underline{Y}} \oplus \left(\mathcal{O}_{\underline{X}} \otimes_{\mathbb{Z}} \mathcal{M}^{\mathrm{gp}}\right)\right) / \langle (\mathrm{d}(\alpha(a)), -\alpha(a) \otimes [a]), (0, 1 \otimes [f^\times(b)]) \rangle_{a \in \mathcal{M}, b \in f^*\mathcal{N}}.$$

The notation $^{\mathrm{gp}}$ indicates that we are taking a (sheafified) Grothendieck group (with $[-]$ denoting the image in the Grothendieck group). Here, we are to understand an element $[a] \in \mathcal{M}^{\mathrm{gp}}$ denoting in the above formula the log differential $\mathrm{dlog}(a)$. This means we can understand the first type of relations, given by quotienting out $(\mathrm{d}(\alpha(a)), -\alpha(a) \otimes [a])$, as corresponding to the formula $\mathrm{d}(\alpha(a)) = \alpha(a)\mathrm{dlog}(a)$. We take the Grothendieck group of $\mathcal{M}$ because we expect dlog to factor through $\mathcal{M}^{\mathrm{gp}}$ (because of the symbolic relation $-\log(a) = \log({}^{1}/_{a})$),

[56] The "associated log-structure" functor is left adjoint to the inclusion functor from pre-log structures to log-structures, where a pre-log structure does not require the condition on $\alpha^{-1}\mathcal{O}_X$.
[57] That is, pairs (d, dlog), with d a derivation and dlog a log derivation over d.

and because we need a bona-fide abelian group in the formula to define a sheaf of $\mathcal{O}_{\underline{X}}$-modules.

The upshot of these considerations is that we can extend methods applicable to smooth schemes in a more general context. We recall first the characterisation of smooth morphisms by the infinitesimal lifting property (see for instance [EGA IV$_4$, Définition 17.1.1]). A morphism of schemes $f\colon X \to Y$ is said to be formally smooth (respectively formally étale, formally unramified) if for every commutative diagram of schemes

$$\begin{tikzcd} U \arrow[r] \arrow[d, hook, "j"'] & X \arrow[d, "f"] \\ T \arrow[r] & Y \end{tikzcd}$$

where $j\colon U \hookrightarrow T$ is a nilpotent infinitesimal thickening (meaning that j is defined by a nilpotent ideal[58]) of *affine* schemes, there exists a morphism $T \to X$ (respectively a unique such morphism, at most one such morphism) making the following diagram commute

$$\begin{tikzcd} U \arrow[r] \arrow[d, hook, "j"'] & X \arrow[d, "f"] \\ T \arrow[r] \arrow[ur, dashed] & Y \end{tikzcd}$$

A morphism of schemes is then smooth (respectively étale, unramified) if and only if it is formally smooth (respectively formally étale, formally unramified) and locally of finite presentation; this is [EGA IV$_4$, Corollaire 17.5.2]. This means we can think of smooth morphisms as those that allow us to lift tangent vectors (and higher-order jets), and similarly for étale or unramified morphisms.
We can then mimic these definitions in the setting of log schemes: a morphism of log schemes is said to be formally log smooth (respectively formally log étale, formally log unramified) if there exist lifts (respectively there exist unique lifts, there exist at most one lift) with respect to morphisms of affine log schemes $j\colon U \hookrightarrow T$ which are *strict* nilpotent closed immersions. Strictness of j is the condition that $j^{\times}\colon j^*\mathcal{M}_T \to \mathcal{M}_U$ be an isomorphism; we don't want to allow j to introduce additional logarithmic data. The finiteness condition to ask for is a bit more subtle: naturally we have to require that the underlying morphism of schemes is locally of finite presentation, but we also need to impose finiteness conditions on the level of log structures.
First we consider the notion of a **chart** for a log structure $\mathcal{M}$ on a scheme X. This consists of an étale cover $\{X_i\}$ of X, monoids M_i and multiplicative homomorphisms $\beta_i\colon \widetilde{M}_i \to \mathcal{M}|_{X_i}$ (where $\widetilde{M}_i$ denotes the sheafification of the constant presheaf M_i) such that the morphisms $(\beta_i)^{\mathrm{a}}\colon (\widetilde{M}_i)^{\mathrm{a}} \longrightarrow \mathcal{M}|_{X_i}$, from the associated log structures $(\widetilde{M}_i)^{\mathrm{a}}$ of $\widetilde{M}_i$ to $\mathcal{M}|_{X_i}$, are isomorphisms.
Consider then the following definitions, for a log structure $\mathcal{M}$ on X:

- $\mathcal{M}$ is **coherent** if it admits a chart of finitely generated monoids,
- $\mathcal{M}$ is **integral** if it is a sheaf of integral monoids,
- $\mathcal{M}$ is **fine** if it is both coherent and integral.

This provides enough of a finiteness condition for reasonable definitions; we then say a morphism of fine log schemes is log smooth (respectively log étale, log unramified) if it is formally log smooth (respectively, formally log étale, formally log unramified) and locally of finite presentation.

We end this section with some much needed examples of log schemes.
Let R be a discrete valuation ring. Then there are three natural log structures on $S = \mathrm{Spec}(R)$:

- the trivial log structure,
- the canonical log structure, which is the log structure attached to $\mathbb{N} \to R$, $1 \mapsto \varpi$ for some uniformiser ϖ of R,

[58] In fact it suffices to check the criterion on square zero ideals, as we can successively extend the lifting level-by-level only using ideals of square zero [EGA IV$_4$, Remarques 17.1.2, (ii)].

- the hollow log structure, attached to $\mathbb{N} \to R$, $1 \mapsto 0$.

The first log structure corresponds to doing classical (non-logarithmic) geometry; the second to working with semistable schemes over S; the third to working with special fibres of semistable schemes over S.
The crucial example is that of semistable schemes, of course. Given $f\colon \mathcal{X} \to S$ a semistable scheme, we obtain a fine log structure. Then, when $\mathcal{X}$ and S are both equipped with the canonical log structures, $f\colon \mathcal{X} \to S$ is log smooth.

9.8 The log-crystalline site

With these basic notions of log geometry, we can hope to adjust the definition of the crystalline topos of section 5.3 to account for the case of non-smooth, semistable reduction. This leads us to considering log PD-schemes, which are simply log schemes whose underlying scheme is equipped with a PD-structure. The notion of PD-thickening then needs only the change that the closed embedding must be **strict** ($\iota^\times$ must be an isomorphism).

Starting with a fine log PD-scheme Y/S, we can define the (étale) **log-crystalline site** Log-Cris(Y/S) analogously to the crystalline site of section 5.3. We have to use the étale topology, rather than the Zariski topology, because of the involvement of log structures. An object of the log-crystalline site is given by the data of an étale scheme U over $\underline{Y}$, a fine log scheme T over S, and a (log) PD-thickening $\iota\colon U \hookrightarrow T$ of U (with the log and PD-structures on U inherited from X). Morphisms of such objects are the usual commutative diagrams. The coverings are now étale covers: the Grothendieck topology is generated by families of morphisms $(\iota_i\colon U_i \hookrightarrow T_i) \to (\iota\colon U \hookrightarrow T)$ such that $T_i \to T$ is a jointly surjective family of étale morphisms; moreover we require that these morphisms induce isomorphisms $U_i \cong T_i \times_T U$. The corresponding **log-crystalline topos** is then written $(X/S)_{\text{log-cris}}$. As per usual, we have a structure sheaf $\mathcal{O}_{Y/S}\colon (\iota\colon U \hookrightarrow T) \mapsto \Gamma(T, \mathcal{O}_T)$, and also as previously sheaves $\mathbb{G}_a$, $\mathcal{J}$. We now also have the sheaf of monoids $\mathcal{M}_{Y/S}\colon (\iota\colon U \hookrightarrow T) \mapsto \Gamma(T, \mathcal{M}_T)$. The intrinsic cohomology of this topos is log-crystalline cohomology.

As for the natural notion of coefficients, we have a similar description as in section 5.3.1, with sheaves on the log-crystalline site being given by compatible systems of étale sheaves; we can also define crystals of $\mathcal{O}_{Y/S}$-modules in the same way. We then have equivalences, when $f\colon Y \to S$ is log smooth, between crystals, hyper PD-stratifications and flat quasi-nilpotent connections [K. Kato. Log structures, Theorem 6.2, 6.7].

Just as Theorem 5.3 related crystalline and de Rham cohomologies in the good reduction case, we have an analogous result in the log-crystalline context.
To set the stage, consider a proper scheme $f\colon \mathcal{X} \to \mathcal{O}_K$ with semistable reduction, special fibre Y/κ, and smooth generic fibre X/K. Just as we were working with the lone special fibre Y/κ and Y/W in the crystalline context, here we work with $Y^\curlywedge/\kappa^\circ$ and $Y^\curlywedge/W^\circ$. The notational difference accentuates that we are working not only with the schematic special fibre Y, but we are also remembering its log structure $\mathcal{M}_{Y^\curlywedge}$ inherited from $\mathcal{X}$. Here κ° means that κ is equipped with the hollow log structure, which we also lift to the Witt vectors, yielding W_n°. This is the natural context for studying the log-crystalline cohomology, as opposed to, for instance, considering the canonical log structure $W^\curlywedge$ on W, because we want to be considering Y as a W-scheme with empty generic fibre. On the other hand, were we to work with a lift of Y to W, we would want to use the canonical log structure on W, not the hollow one. Note that $Y^\curlywedge$, as a log scheme, is *not* determined by the underlying schematic special fibre Y: the log structure also captures some second-order information.
The comparison between crystalline and de Rham cohomologies in this context is as follows [Hyodo–Kato, Theorem 5.1] [Tsuji. C_{st}, Proposition 4.4.9]

Theorem 9.3 (Hyodo, Kato) — Dependent on a choice of uniformiser ϖ of $\mathcal{O}_K$, there is an isomorphism of K-vector spaces

$$\rho_\varpi\colon \mathrm{H}^i_{\mathrm{cris}}\left(Y^\curlywedge/W^\circ, \mathcal{O}_{Y^\curlywedge/W^\circ}\right) \otimes_W K \xrightarrow{\sim} \mathrm{H}^i_{\mathrm{dR}}(X/K).$$

In addition, we have the formula $\rho_{\varpi u} = \rho_\varpi \circ \exp(\log(u)N)$ for any unit $u \in \mathcal{O}_K^\times$, where N is the logarithm of monodromy endomorphism on the crystalline cohomology that we will construct in section 9.9. In particular, the endomorphism $\rho_\varpi \circ N \circ \rho_\varpi^{-1}$ on $\mathrm{H}^i_{\mathrm{dR}}(X/K)$ is well-defined independent of choices, and matches up with the logarithm of the monodromy of the Gauss–Manin connection defined on $\mathbf{R}^i f_* \Omega^\bullet_{X/K}$. ◇□

9.9 Monodromy and weights in log-crystalline cohomology

We want to introduce considerations of weights, and define a logarithm of monodromy operator N, in this new context. To start off then, we want to give the log-crystalline analogue of the weight filtration on $\Omega^\bullet_{X/K}(\log D)$ that we saw in section 9.6.

We saw (Theorem 5.5) that the correct object in crystalline cohomology that replaces the de Rham complex $\Omega^\bullet_{X/K}$ is the de Rham–Witt complex $\mathcal{W}_\star\Omega^\bullet_{X/K}$. In the log-crystalline case, Hyodo and Kato define the de Rham–Witt complex for log schemes, which serves the same purpose as the complex of log differentials $\Omega^\bullet_{X/K}(\log D)$ in this new setting. In particular, its hypercohomology computes the log-crystalline cohomology. This will allow us to study the Gauss–Manin connection in this context, and define a weight spectral sequence.

To define a connection on log-crystalline cohomology, we want an analogue of the exact sequence (9.6.1)

$$0 \longrightarrow \Omega^\bullet_{X/S}(\log D)[-1] \xrightarrow{\alpha \mapsto \mathrm{dlog}(z)\wedge\alpha} \Omega^\bullet_{X/\mathbb{C}}(\log D) \longrightarrow \Omega^\bullet_{X/S}(\log D) \longrightarrow 0,$$

which allowed us to define a connection on the de Rham cohomology of a semistable family $f\colon X \to S$ of complex varieties over the unit disc, with $D = X_0$ a divisor with normal crossings on X.

The way to do this is use a logarithmic de Rham–Witt complex $\mathcal{W}_\star\Omega^\bullet_{Y^\curlywedge/\kappa^\circ}$ whose hypercohomology computes the log-crystalline cohomology $\mathrm{H}^\bullet_{\mathrm{cris}}(Y^\curlywedge/W^\circ, \mathcal{O}_{Y^\curlywedge/W^\circ})$, as defined by Hyodo and Kato. This complex will play the rôle of $\Omega^\bullet_{X/S}(\log D)_{\{0\}}$ in this context. As for $\Omega^\bullet_{X/\mathbb{C}}(\log D)_{\{0\}}$, for lack of a better option we can construct explicitly an analogue, which we will denote by $\mathcal{W}_\star\widetilde{\Omega}^\bullet_{Y^\curlywedge}$. We begin with the following explicit description/definition of $\mathcal{W}_n\Omega^i_{Y^\curlywedge/\kappa^\circ}$ given by [Hyodo–Kato, Proposition 4.6] and corrected by [Nakkajima. ${}_W\mathrm{E}_1^{p,q}$, §7]:

$$\mathcal{W}_n\Omega^i_{Y^\curlywedge/\kappa^\circ} \cong \left(\left(\mathcal{W}_n\mathcal{O}_Y \otimes_{\mathbb{Z}} \bigwedge^i \mathcal{M}_Y^{\mathrm{gp}}/f^{-1}(\mathcal{M}_\kappa^{\mathrm{gp}})\right) \oplus \left(\mathcal{W}_n\mathcal{O}_Y \otimes_{\mathbb{Z}} \bigwedge^{i-1} \mathcal{M}_Y^{\mathrm{gp}}/f^{-1}(\mathcal{M}_\kappa^{\mathrm{gp}})\right)\right) \Big/ (\mathcal{F} + \mathcal{G}),$$

where $\mathcal{F}$, $\mathcal{G}$ are the subsheaves spanned by local sections of the form:

$$\mathcal{F}: \qquad (\varepsilon_j(\alpha(a_1)) \otimes (a_1 \wedge \cdots \wedge a_i), 0) - p^j\big(0, \varepsilon_j(\alpha(a_1)) \otimes (a_2 \wedge \cdots \wedge a_i)\big),$$

for $a_1, \ldots, a_i$ local sections of $\mathcal{M}_Y$, $j \in \mathbb{N}$, and where $\varepsilon_j\colon \mathcal{O}_Y \to \mathcal{W}_n\mathcal{O}_Y$ is the morphism locally given by $\varepsilon_j\colon b \mapsto (\underbrace{0, \ldots, 0}_{j}, b, 0, \ldots)$, and

$$\mathcal{G}: \qquad \big(0, \varepsilon_j(\alpha(a_2)) \otimes (a_2 \wedge \cdots \wedge a_i)\big),$$

for $a_2, \ldots, a_i$ local sections of $\mathcal{M}_Y$. In both cases we have omitted the notation $[-]$ to denote the image in the Grothendieck group $\mathcal{M}^{\mathrm{gp}}$.

For this to make any sense, we want the hypercohomology of $\mathcal{W}_\star\Omega^\bullet_{Y^\curlywedge/\kappa^\circ}$ to compute the log-crystalline cohomology, generalising Theorem 5.5.

Theorem 9.4 (Hyodo, Kato, Nakkajima) — There are compatible (as n varies) canonical isomorphisms of complexes of sheaves on the étale site of Y

$$\mathrm{R}(\mathrm{u}_{Y^\curlywedge/W_n^\circ})_*(\mathcal{O}_{Y^\curlywedge/W_n^\circ}) \xrightarrow{\sim} \mathcal{W}_n\Omega^\bullet_{Y^\curlywedge/\kappa^\circ}. \qquad \diamond$$

Proof: See [Nakkajima. ${}_W\mathrm{E}_1^{p,q}$, Theorem 7.19], where $\mathcal{W}_n\Omega^\bullet_{Y^\curlywedge/\kappa^\circ}$ is denoted $(\mathcal{W}_n\Lambda^\bullet_Y)''$ (which is itself isomorphic to the complex $\mathcal{W}_n\Lambda^\bullet_Y$ of the quoted theorem, by [Nakkajima. ${}_W\mathrm{E}_1^{p,q}$, Theorem 7.5]). □

It then makes sense to define an absolute version of this logarithmic de Rham–Witt complex as:

$$\mathcal{W}_n\widetilde{\Omega}^i_{Y^\curlywedge} := \left(\left(\mathcal{W}_n\mathcal{O}_Y \otimes_{\mathbb{Z}} \bigwedge^i \mathcal{M}_Y^{\mathrm{gp}}/\alpha^{-1}(\kappa^\times)\right) \oplus \left(\mathcal{W}_n\mathcal{O}_Y \otimes_{\mathbb{Z}} \bigwedge^{i-1} \mathcal{M}_Y^{\mathrm{gp}}/\alpha^{-1}(\kappa^\times)\right)\right) \Big/ (\mathcal{F}+\mathcal{G}).$$

Here $\mathcal{F}$ and $\mathcal{G}$ are defined as above. The only difference is that we are now no longer quotienting out by $f^{-1}(\mathcal{M}_\kappa^{\mathrm{gp}})$ (but instead just by constants), so we can expect it to be a suitable absolute version of the relative log de Rham–Witt sheaves $\mathcal{W}_n\Omega^i_{Y^\curlywedge/\kappa^\circ}$, and so mirror $\Omega^i_{X/\mathbb{C}}(\log D)_{\{0\}}$.
Finally now we can state:

Proposition 9.5 (Nakkajima) — Let τ be a generator of the hollow log structure on κ. The sequence of complexes of sheaves with G_κ-action

$$0 \longrightarrow \mathcal{W}_\star\Omega^\bullet_{Y^\curlywedge/\kappa^\circ}(-1)[-1] \xrightarrow{\alpha\mapsto \mathrm{dlog}(\tau)\wedge\alpha} \mathcal{W}_\star\widetilde{\Omega}^\bullet_{Y^\curlywedge} \longrightarrow \mathcal{W}_\star\Omega^\bullet_{Y^\curlywedge/\kappa^\circ} \longrightarrow 0$$

is exact. ◇

Proof: See [Nakkajima. ${}_{\mathcal{W}}\mathrm{E}_1^{p,q}$, Theorem 11.1], in which the complexes $\mathcal{W}_\star\Omega^\bullet_{Y^\curlywedge/\kappa^\circ}$ and $\mathcal{W}_\star\widetilde{\Omega}^\bullet_{Y^\curlywedge}$ are respectively denoted $\left(\mathcal{W}_\star\Lambda^\bullet_Y\right)''$ and $\left(\mathcal{W}_\star\widetilde{\Lambda}^\bullet_Y\right)''$. □

This allows us finally to define the logarithm of monodromy N on log-crystalline cohomology, as the boundary morphism obtained in hypercohomology from the short exact sequence of the above theorem:

$$N\colon \mathbf{H}^q_{\text{ét}}\left(Y, \mathcal{W}_\star\Omega^\bullet_{Y^\curlywedge/\kappa^\circ}\right) \longrightarrow \mathbf{H}^q_{\text{ét}}\left(Y, \mathcal{W}_\star\Omega^\bullet_{Y^\curlywedge/\kappa^\circ}\right)(-1).$$

As the hypercohomology of the logarithmic de Rham–Witt complex computes log-crystalline cohomology by Theorem 9.4, we obtain[59]

$$N\colon \mathrm{H}^i_{\mathrm{cris}}\left(Y^\curlywedge/W^\circ, \mathcal{O}_{Y^\curlywedge/W^\circ}\right) \to \mathrm{H}^i_{\mathrm{cris}}\left(Y^\curlywedge/W^\circ, \mathcal{O}_{Y^\curlywedge/W^\circ}\right)(-1). \tag{9.9.1}$$

The Tate twist indicates that N is not quite compatible with the Frobenius endomorphism, but is compatible up to a twist: $N\varphi = p\varphi N$, as $\varphi_{M(-1)} = p\varphi_M$. This corresponds to the factor of $2\pi\mathrm{i}$ in the formula $N = 2\pi\mathrm{i}\,\mathrm{res}_0\nabla$ of section 9.6.
We refer to [Nakkajima. ${}_{\mathcal{W}}\mathrm{E}_1^{p,q}$, Theorem 11.4] for the proof that this definition of N agrees with other possible definitions of N.

9.10 The ring $\mathbf{B}_{\mathrm{st}}^+$

Let's now go back to our project of p-adic integration. We started off considering $\int_\gamma {}^{\mathrm{d}z}/_z$ as giving the monodromy of the logarithm, and defined $\mathrm{t} = \log(\widetilde{\varepsilon})$ as a p-adic $2\pi\mathrm{i}$, a period for the p-adic cyclotomic character.
However, we saw that when we have a singular special fibre, there can be other periods to consider too. In particular, the Picard–Lefschetz formula leads us to desire a period for the tame Kummer character t_p. For instance, the action of inertia on an absolute cycle of the Tate curve E_p is given by $g\cdot\eta = \eta + \mathrm{c}_p(g)\nu$. In terms then of the integral of the invariant differential ω over η, we expect the integral $\mathrm{u} = \int_\eta \omega$ to satisfy

$$g\cdot\mathrm{u} = \mathrm{u} + \mathrm{c}_p(g)\mathrm{t}, \tag{9.10.1}$$

as g acts on ν through the cyclotomic character, so that the integral over ν is (up to a constant, which we'll ignore) just $\mathrm{t} = \log(\widetilde{\varepsilon})$.
Now, whereas χ_p captures the action of the Galois group on p-power roots of unity, t_p (or c_p) captures the action on p-power roots of p. So, on top of a system $\widetilde{\varepsilon} = [(\varepsilon_0, \varepsilon_1, \varepsilon_2, \ldots)] \in \mathbf{A}_{\mathrm{inf}}$ of p-power roots of unity ($\varepsilon_0 = 1$, ε_1 a primitive p-th root of unity, $\varepsilon_n^p = \varepsilon_{n-1}$), we want $\widetilde{p} = [p^\flat] = [(p^{(0)}, p^{(1)}, p^{(2)}, \ldots)] \in \mathbf{A}_{\mathrm{inf}}$ satisfying $p^{(0)} = p$ and $(p^{(n)})^p = p^{(n-1)}$, a compatible system of p-power roots of p. Formally writing $\mathrm{u} = \log(\widetilde{p})$ for the moment, we can see that $g\cdot\mathrm{u} = \log(g\cdot\widetilde{p}) = \log\left(\widetilde{\varepsilon}^{\,\mathrm{c}_p(g)}\widetilde{p}\right) = \log(\widetilde{p}) + \mathrm{c}_p(g)\log(\widetilde{\varepsilon}) = \mathrm{u} + \mathrm{c}_p(g)\mathrm{t}$. This means that, just as

[59] It is however not completely straightforward to verify that N yields an endomorphism on the log-crystalline cohomology. See [Hyodo–Kato, 3.6] and [Nakkajima. ${}_{\mathcal{W}}\mathrm{E}_1^{p,q}$, § 11].

$t = \log(\widetilde{\varepsilon})$ was a period for the cyclotomic character, we can consider $\mathfrak{u} = \log(\widetilde{p})$ as a period for the fundamental tame character t_p, via the above formula.

We go back to the definition now of $\log(\widetilde{p})$, and in particular to considering the convergence of the series that defines $\log(\widetilde{p})$.
Considering the description of $\mathbf{B}_{\mathrm{dR}}^+$ as the $\ker(\theta)$-adic completion of $\mathbf{A}_{\mathrm{inf}}$, for $\log(\widetilde{p})$ to converge we want $1-\widetilde{p}$ to be in the kernel of θ. However $\theta(1-\widetilde{p}) = 1-p$, so this doesn't quite work. Instead, let's start by defining $\log(p)$ separately (as an element of $\mathbb{C}_p$) and $\log(\widetilde{p}/p)$, which we can do without trouble as $\theta(1-\widetilde{p}/p) = 0$. We have:

$$\log\left(\frac{\widetilde{p}}{p}\right) = -\sum_{n=1}^{\infty} \frac{\left(1-\frac{\widetilde{p}}{p}\right)^n}{n},$$
$$\log(\widetilde{p}) := \log\left(\frac{\widetilde{p}}{p}\right) + \log(p).$$

Thus $\log(\widetilde{p})$ is an element of $\mathbf{B}_{\mathrm{dR}}^+$. However the series defining $\log(\widetilde{p})$ *does not converge* in $\mathbf{B}_{\mathrm{cris}}^+$, and in fact $\log(\widetilde{p})$ is transcendental over $\mathbf{B}_{\mathrm{cris}}^+$. This allows us to define $\mathbf{B}_{\mathrm{st}}^+ = \mathbf{B}_{\mathrm{cris}}[\mathfrak{u}]$, with $\mathfrak{u}$ a formal variable. A choice of $\log(p)$ then yields an embedding $\mathbf{B}_{\mathrm{st}}^+ \hookrightarrow \mathbf{B}_{\mathrm{dR}}^+$, $\mathfrak{u} \mapsto \log(\widetilde{p})$.[60] Note however that there is no natural embedding independent of choices. We also define $\mathbf{B}_{\mathrm{st}} = \mathbf{B}_{\mathrm{st}}^+[1/t]$, this is Fontaine's **ring of semistable periods**.

What structures do we have then on $\mathbf{B}_{\mathrm{st}}$? We have the Galois action of G_K inherited from $\mathbf{B}_{\mathrm{cris}}$ (with G_K acting on $\mathfrak{u}$ through formula (9.10.1)), the Frobenius operator inherited from $\mathbf{B}_{\mathrm{cris}}$, which we extend in the obvious manner to $\mathbf{B}_{\mathrm{st}}$ via $\varphi(\mathfrak{u}) = p\mathfrak{u}$ (we have in mind: $\varphi(\log(\widetilde{p})) = \log(\widetilde{p}^p) = p\log(\widetilde{p})$). We know that $(\mathbf{B}_{\mathrm{cris}})^{G_K} = K_0$, so that also $(\mathbf{B}_{\mathrm{st}})^{G_K} = K_0$. The embedding $\mathbf{B}_{\mathrm{cris}} \otimes_{K_0} K \to \mathbf{B}_{\mathrm{dR}}$ then extends to an embedding $\mathbf{B}_{\mathrm{st}} \otimes_{K_0} K \to \mathbf{B}_{\mathrm{dR}}$ once $\log(p)$ is chosen. We will study $(\mathbf{B}_{\mathrm{st}})^{\varphi=1}$ below, using the fundamental exact sequence of $\mathbf{B}_{\mathrm{cris}}$ that we covered in section 5.5.
Next, we have the Hodge filtration $F^\bullet$, inherited from the Hodge filtration on $\mathbf{B}_{\mathrm{dR}}$ once an embedding $\mathbf{B}_{\mathrm{st}} \hookrightarrow \mathbf{B}_{\mathrm{dR}}$ is chosen.
In additional to all this, we also expect there to be endomorphism N that mirrors the Picard–Lefschetz formula. It should be $\mathbf{B}_{\mathrm{cris}}$-linear (as there is no monodromy in the case of good reduction). It should also be a derivation, as it is to be considered as the logarithm of a monodromy automorphism.[61] Thus it suffices to define $N(\mathfrak{u})$; mirroring the Picard–Lefschetz formula and the expression $\rho(g) = \exp(N t_\ell(g))$, it makes sense to choose $N(\mathfrak{u}) = -1$, with the sign supposed to reflect the sign that appears in the Picard–Lefcshetz formula in the case of relative dimension 1.
How do these all relate? For a start the Galois action of G_K commutes with everything else: φ, N and $F^\bullet$. Also $F^\bullet$ commutes with both φ and N, however φ and N do *not* commute. We have the following relation that parallels the case of $\ell \neq p$, and mirrors the relation in log-crystalline cohomology:

$$N\varphi = p\varphi N.$$

Indeed, it suffices to note that $N\varphi\mathfrak{u} = Np\mathfrak{u} = -p = p\varphi N\mathfrak{u}$, as N is $\mathbf{B}_{\mathrm{cris}}$-linear, with $N\mathfrak{u} = -1$, and φ is multiplicative and $\mathbb{Q}_p$-linear. Another way to state this relation is that N commutes with φ when considered as a morphism $N\colon \mathbf{B}_{\mathrm{st}} \to \mathbf{B}_{\mathrm{st}}(-1)$.

Finally, we can apply the results of section 5.5 to elucidate the interaction of all these extra structures on the

[60] We could have also performed the above construction with an arbitrary $q \in \mathfrak{m}_K$ instead of p, so we can more generally think of $\mathfrak{u}$ as denoting $\log(\widetilde{q})$.
[61] The formula $T(ab) = T(a)T(b)$ implies $\log(T)(ab) = \log(T)(a)b + a\log(T)(b)$.

K_0-algebra $\mathbf{B}_{\mathrm{st}}$. We want to adapt the double complex

$$\begin{tikzcd}
\mathbb{Q}_p \arrow[r] \arrow[d] & \mathbf{B}_{\mathrm{cris}} \arrow[r, "\varphi-1"] \arrow[d] & \mathbf{B}_{\mathrm{cris}} \\
\mathbf{B}^+_{\mathrm{dR}} \arrow[r] & \mathbf{B}_{\mathrm{dR}} &
\end{tikzcd}$$

to the situation of $\mathbf{B}_{\mathrm{st}}$. We simply need to add N with the relation $N\varphi = p\varphi N$. Choosing an embedding $\mathbf{B}_{\mathrm{st}} \hookrightarrow \mathbf{B}_{\mathrm{dR}}$, this leads to the triple complex:

$$\begin{tikzcd}
\mathbb{Q}_p \arrow[rr] \arrow[dd] & & \mathbf{B}_{\mathrm{st}} \arrow[rr, "\varphi-1"] \arrow[dl, "N"'] \arrow[dd] & & \mathbf{B}_{\mathrm{st}} \arrow[dl, "N"] \\
& \mathbf{B}_{\mathrm{st}} \arrow[rr, "1-p\varphi"] & & \mathbf{B}_{\mathrm{st}} & \\
\mathbf{B}^+_{\mathrm{dR}} \arrow[rr] & & \mathbf{B}_{\mathrm{dR}} & &
\end{tikzcd}$$

The notations $\mathcal{K}_{\mathrm{cris}}$, $\mathcal{K}_{\mathrm{st}}$ will be justified in section 16, where these complexes will feature prominently.

Generalising Corollary 5.7, we see that the total complex associated with the above triple complex is exact. In particular, this means that we have $F^0(\mathbf{B}_{\mathrm{st}})^{\varphi=1,N=0} = \mathbb{Q}_p$. As for the behaviour of N, we have the following exact sequence:

$$0 \longrightarrow F^0\mathbf{B}_{\mathrm{cris}} \to F^0\mathbf{B}_{\mathrm{st}} \xrightarrow{N} \mathbf{B}_{\mathrm{st}} \longrightarrow 0.$$

This follows immediately from the description $\mathbf{B}_{\mathrm{st}} \cong \mathbf{B}_{\mathrm{cris}}[\mathrm{u}]$ with $N = -\frac{\mathrm{d}}{\mathrm{du}}$, once $\log(p)$ is chosen.

Finally, we want to make sure that $\mathbf{B}_{\mathrm{st}}$ is an admissible $(\mathbb{Q}_p, \mathrm{G}_K)$-regular ring. This follows without difficulty from the proof of Theorem 8A.9 which covered the $(\mathbb{Q}_p, \mathrm{G}_K)$-regularity of $\mathbf{B}_{\mathrm{cris}}$, as we simply need to replace the observation $(\mathbf{B}_{\mathrm{cris}})^{\mathrm{G}_L} = L_0$ with $(\mathbf{B}_{\mathrm{st}})^{\mathrm{G}_L} = L_0$ to conclude the analogous proof.

10 The semistable comparison theorem

We are now ready to state the enhancement to the semistable situation of the crystalline comparison theorem (Theorem 6.1). The statement is as expected: whereas with $\mathbf{B}_{\mathrm{cris}}$ we only picked up periods of varieties with good reduction, we can hope to use $\mathbf{B}_{\mathrm{st}}$ to also detect periods of varieties with semistable reduction. The following result was conjectured by Fontaine and Jannsen.

Theorem 10.1 (Tsuji, Faltings, Niziol) — Let X be a smooth proper variety over K which extends to a semistable model $\mathcal{X}/\mathcal{O}_K$, with special fibre Y. There is a natural functorial comparison isomorphism

$$\mathrm{C}_{\mathrm{st}}\colon \mathrm{H}^i_{\text{ét}}\left(X_{\overline{K}}, \mathbb{Z}_p\right) \otimes_{\mathbb{Z}_p} \mathbf{B}_{\mathrm{st}} \xrightarrow{\sim} \mathrm{H}^i_{\mathrm{cris}}\left(Y^{\curlywedge}/W^\circ, \mathcal{O}_{Y^{\curlywedge}/W^\circ}\right) \otimes_W \mathbf{B}_{\mathrm{st}}.$$

Once again, this isomorphism is compatible with the G_K action, the Frobenius endomorphisms, the filtrations (after $-\otimes_{K_0} K$), the (logarithm of) monodromy endomorphisms N, Poincaré duality, the Künneth formula, Tate twists, the formation of Chern classes of vector bundles, cycle class maps, and is compatible with the associated E_∞-algebra structures implicit on both sides. ◇□

The notation $Y^{\curlywedge}/W^\circ$ emphasises that the right hand side does not solely depend on the special fibre Y, as the log structure on Y is induced from $\mathcal{X}$. This is an essential difference with the crystalline situation: in the crystalline case the cohomology only depended on the special fibre, whereas here we need some small amount of information from $\mathcal{X}$. This is precisely captured by the log structure of $Y^{\curlywedge}$, which as mentioned previously is determined by the reduction modulo $\mathfrak{m}^2$ of $\mathcal{X}$.

We now indicate the relation between the de Rham comparison theorem (Theorem 4.1) and the semistable comparison theorem.
Recall Deligne's method for furnishing the cohomology of an arbitrary smooth scheme X of finite type over $\mathbb{C}$ with a mixed Hodge structure, as we saw in section 9.6: we used a Nagata compactification together with Hironaka's resolution of singularities to embed X in a proper smooth scheme Z, with complement a normal crossings divisor $D \hookrightarrow Z$. The weight and Hodge filtration on the complex of logarithmic differentials $\Omega^\bullet_{Z/\mathbb{C}}(\log D)$ then yielded the requisite mixed Hodge structure on $\mathrm{H}^n_{\mathrm{dR}}(X/\mathbb{C})$.

Now, in the p-adic setting, we are to start with a smooth proper scheme X/K. We aim to prove the de Rham comparison theorem. Trying to reduce the problem to C_{st}, we look to relate X to a semistable scheme. Unfortunately, no result as strong as Hironaka's is available. However, we can get away with asking for slightly less [de Jong. Alterations, Theorem 4.5]:

Theorem 10.2 (de Jong) — Let $\mathcal{X}/\mathcal{O}_K$ be an integral separated scheme which is flat and of finite type over $\mathcal{O}_K$, with smooth geometric fibre X and special fibre Y. There exist:

- a finite extension L/K,
- a smooth scheme $\mathcal{V}/\mathcal{O}_L$,
- a proper, surjective morphism $a\colon \mathcal{V} \to \mathcal{X}$,
- an open immersion $j\colon \mathcal{V} \hookrightarrow \mathcal{Z}$ of smooth schemes over $\mathcal{O}_L$

such that:

- there exists a non-empty open subscheme $\mathcal{U}$ of $\mathcal{X}$ such that the morphism $a^{-1}(\mathcal{U}) \to \mathcal{U}$ induced by a is finite (we call a an **alteration**),
- $\mathcal{Z}$ is projective and strictly semistable over $\mathcal{O}_L$, with geometrically irreducible generic fibre Z,
- $a^{-1}(Y)_{\mathrm{red}} \cup (\mathcal{Z} \smallsetminus j(\mathcal{V}))$ is a strict normal crossings divisor on $\mathcal{Z}$,
- letting $\{\mathcal{C}_i\}_i$ be the irreducible components of the flat part of $\mathcal{Z} \smallsetminus j(\mathcal{V})$, each $\cap_{i\in I}\mathcal{C}_i$ is a union of strictly semistable schemes over $\mathcal{O}_L$. ◇□

Following [Tsuji. C_{st} (survey), Appendix], we can employ alterations to deduce C_{dR} from C_{st}. Starting with an arbitrary smooth proper scheme X/K, we choose a flat proper integral model $\mathcal{X}/\mathcal{O}_K$. Passing to a finite extension and picking a connected component, we may assume that $\mathcal{X}$ is geometrically connected. We use the above theorem to obtain a finite extension L/K, a semistable scheme $\mathcal{V}/\mathcal{O}_L$, and an alteration $a\colon \mathcal{V} \to \mathcal{X}$. Moreover, by properness of $\mathcal{X}/\mathcal{O}_K$, we may assume that $\mathcal{V}$ is proper too.
Applying the semistable comparison theorem to the semistable scheme $\mathcal{V}/\mathcal{O}_L$, with special fibre $V = \mathcal{V}_L$, we obtain:

$$\mathrm{C}_{\mathrm{st}}(\mathcal{V}/\mathcal{O}_L)\colon \mathrm{H}^i_{\text{ét}}\left(V_{\overline{K}}, \mathbb{Q}_p\right) \otimes_{\mathbb{Q}_p} \mathbf{B}_{\mathrm{st}} \xrightarrow{\sim} \mathrm{H}^i_{\mathrm{cris}}\left(\mathcal{V}^{\curlywedge}_{\kappa_L}/\mathcal{W}(\kappa_L)^\circ, \mathcal{O}_{\mathcal{V}^{\curlywedge}_{\kappa_L}/\mathcal{W}(\kappa_L)^\circ}\right) \otimes_{\mathcal{W}(\kappa_L)} \mathbf{B}_{\mathrm{st}}.$$

Heading for the de Rham comparison theorem, we apply $- \otimes_{\mathbf{B}_{\mathrm{st}}} \mathbf{B}_{\mathrm{dR}}$ to both sides. We can thus forget about most of the extra structure on the right hand side, and only remember the filtration on $\mathrm{H}^i_{\mathrm{cris}}\left(\mathcal{V}^{\curlywedge}_{\kappa_L}/\mathcal{W}(\kappa_L)^\circ, \mathcal{O}_{\mathcal{V}^{\curlywedge}_{\kappa_L}/\mathcal{W}(\kappa_L)^\circ}\right) \otimes_{\mathcal{W}(\kappa_L)} L$, which is just given by the Hodge filtration on $\mathrm{H}^i_{\mathrm{dR}}(V/L)$ through the isomorphism of Theorem 9.3. We thus obtain a comparison isomorphism

$$\mathrm{C}_{\mathrm{dR}}(V/L)\colon \mathrm{H}^i_{\text{ét}}\left(V_{\overline{K}}, \mathbb{Q}_p\right) \otimes_{\mathbb{Q}_p} \mathbf{B}_{\mathrm{dR}} \xrightarrow{\sim} \mathrm{H}^i_{\mathrm{dR}}(V/L) \otimes_L \mathbf{B}_{\mathrm{dR}}$$

which is compatible with all structures at hand.
It then remains to relate $\mathrm{H}^i_{\text{ét}}\left(V_{\overline{K}}, \mathbb{Q}_p\right)$ with $\mathrm{H}^i_{\text{ét}}\left(X_{\overline{K}}, \mathbb{Q}_p\right)$, and $\mathrm{H}^i_{\mathrm{dR}}(V/L)$ with $\mathrm{H}^i_{\mathrm{dR}}(X/K)$. Note that we are now only dealing with the generic fibres V and X.

We must use our alteration $a\colon \mathcal{V} \to \mathcal{X}$. Passing to generic fibres over L, a induces a generically finite and flat morphism $\tilde{a}\colon V \to X_L$, between smooth schemes of equal dimensions. As a consequence, on the étale side we acquire a trace map tr: $\mathrm{R}\tilde{a}_*\mathbb{Q}_p \to \mathbb{Q}_p$ which when composed with the canonical morphism $\mathbb{Q}_p \to \mathrm{R}\tilde{a}_*\mathbb{Q}_p$ yields multiplication by d, where d is the degree of $\tilde{a}$; and similarly on the de Rham side. This then exhibits $\mathrm{H}^i_{\text{ét}}(X_{\bar{K}}, \mathbb{Q}_p)$ and $\mathrm{H}^i_{\mathrm{dR}}(X_L/L)$ as matching direct summands of $\mathrm{H}^i_{\text{ét}}(V_{\bar{K}}, \mathbb{Q}_p)$ and $\mathrm{H}^i_{\mathrm{dR}}(V/L)$. As a consequence, we obtain a commutative diagram

$$\begin{tikzcd}
\mathrm{H}^i_{\text{ét}}(X_{\bar{K}}, \mathbb{Q}_p) \otimes_{\mathbb{Q}_p} \mathbf{B}_{\mathrm{dR}} \arrow[r, "a^*\mathrm{C}_{\mathrm{dR}}(V/L)", "\sim"'] \arrow[d, hook, "a^*\otimes 1"'] & \mathrm{H}^i_{\mathrm{dR}}(X/K) \otimes_K \mathbf{B}_{\mathrm{dR}} \arrow[d, hook, "a^*\otimes 1"] \\
\mathrm{H}^i_{\text{ét}}(V_{\bar{K}}, \mathbb{Q}_p) \otimes_{\mathbb{Q}_p} \mathbf{B}_{\mathrm{dR}} \arrow[r, "\mathrm{C}_{\mathrm{dR}}(V/L)", "\sim"'] & \mathrm{H}^i_{\mathrm{dR}}(V/L) \otimes_L \mathbf{B}_{\mathrm{dR}},
\end{tikzcd}$$

where the horizontal arrows are isomorphisms which commute with the action of G_L, and are compatible with the Hodge filtrations.

We then want to show that the isomorphism of the top row is in fact G_K-equivariant, and not just G_L-equivariant. To do so, we need the following result:

Lemma 10.3 — The comparison isomorphism $a^*\mathrm{C}_{\mathrm{dR}}(V/L)$ depends only on X. In other words, were we to choose different $\mathcal{X}'/\mathcal{O}_K$, L'/K, $\mathcal{V}'/\mathcal{O}_L$, $a'\colon \mathcal{V}' \to \mathcal{X}'$, there would be an equality of morphisms $a^*\mathrm{C}_{\mathrm{dR}}(V/L) = (a')^*\mathrm{C}_{\mathrm{dR}}(V'/L')$. ◇

Proof: See [Tsuji. C_{st} (survey), Proposition A6]. □

Now, given $g \in \mathrm{G}_K$, we consider the twists V^g and $\mathcal{V}^g$, obtained by base-changing V and $\mathcal{V}$ by $g\colon \mathrm{Spec}(g\mathcal{O}_L) \to \mathrm{Spec}(\mathcal{O}_L)$. Write also a^g for the composite $\mathcal{V}^g \xrightarrow{\sim} \mathcal{V} \to \mathcal{X}$; a^g then lies over $\mathcal{O}_K \hookrightarrow g\mathcal{O}_L$. Using both a and a^g then yields the following commutative diagram:

$$\begin{tikzcd}
\mathrm{H}^i_{\text{ét}}(X_{\bar{K}}, \mathbb{Q}_p) \otimes_{\mathbb{Q}_p} \mathbf{B}_{\mathrm{dR}} \arrow[r, "a^*\mathrm{C}_{\mathrm{dR}}(V/L)", "\sim"'] \arrow[d, hook, "a^*\otimes 1"'] & \mathrm{H}^i_{\mathrm{dR}}(X/K) \otimes_K \mathbf{B}_{\mathrm{dR}} \arrow[d, hook, "a^*\otimes 1"] \\
\mathrm{H}^i_{\text{ét}}(V_{\bar{K}}, \mathbb{Q}_p) \otimes_{\mathbb{Q}_p} \mathbf{B}_{\mathrm{dR}} \arrow[r, "\mathrm{C}_{\mathrm{dR}}(V/L)", "\sim"'] \arrow[d, "g^*\otimes g"', "\wr"] & \mathrm{H}^i_{\mathrm{dR}}(V/L) \otimes_L \mathbf{B}_{\mathrm{dR}} \arrow[d, "g^*\otimes g", "\wr"'] \\
\mathrm{H}^i_{\text{ét}}\left(V^g_{\bar{K}}, \mathbb{Q}_p\right) \otimes_{\mathbb{Q}_p} \mathbf{B}_{\mathrm{dR}} \arrow[r, "\mathrm{C}_{\mathrm{dR}}(V^g/gL)", "\sim"'] & \mathrm{H}^i_{\mathrm{dR}}(V^g/gL) \otimes_{gL} \mathbf{B}_{\mathrm{dR}} \\
\mathrm{H}^i_{\text{ét}}(X_{\bar{K}}, \mathbb{Q}_p) \otimes_{\mathbb{Q}_p} \mathbf{B}_{\mathrm{dR}} \arrow[r, "(a^g)^*\mathrm{C}_{\mathrm{dR}}(V^g/gL)", "\sim"'] \arrow[u, hook, "(a^g)^*\otimes 1"] & \mathrm{H}^i_{\mathrm{dR}}(X/K) \otimes_K \mathbf{B}_{\mathrm{dR}}. \arrow[u, hook, "(a^g)^*\otimes 1"']
\end{tikzcd}$$

Applying the previous lemma to $L' = gL$, $\mathcal{V}' = \mathcal{V}^g$, $a' = a^g$, the top and bottom horizontal arrows in the previous commutative diagram are equal. To wrap it all up, we note that $g^* \circ a^* = (a^g)^* \circ g$, where the g on the right hand side corresponds to the natural action by transport of structure (which is trivial on the de Rham side). This proves the commutativity of the diagram

$$\begin{tikzcd}
\mathrm{H}^i_{\text{ét}}(X_{\bar{K}}, \mathbb{Q}_p) \otimes_{\mathbb{Q}_p} \mathbf{B}_{\mathrm{dR}} \arrow[r, "\mathrm{C}_{\mathrm{dR}}", "\sim"'] \arrow[d, "g\otimes g"', "\wr"] & \mathrm{H}^i_{\mathrm{dR}}(X/K) \otimes_K \mathbf{B}_{\mathrm{dR}} \arrow[d, "1\otimes g", "\wr"'] \\
\mathrm{H}^i_{\text{ét}}(X_{\bar{K}}, \mathbb{Q}_p) \otimes_{\mathbb{Q}_p} \mathbf{B}_{\mathrm{dR}} \arrow[r, "\mathrm{C}_{\mathrm{dR}}", "\sim"'] & \mathrm{H}^i_{\mathrm{dR}}(X/K) \otimes_K \mathbf{B}_{\mathrm{dR}},
\end{tikzcd}$$

which is the required equivariance of C_{dR} with respect to the action of G_K. The compatibility of C_{dR} with the other structures then straightforwardly follows from knowing their compatibility with C_{st} [Tsuji. C_{st} (survey), Appendix, p.368].

11 Filtered (φ, N)-modules

It is again useful to axiomatise the various structures on the log-crystalline cohomology, just as we did for filtered φ-modules in section 7. We only need to add an endomorphism N, to be thought of as the logarithm of a monodromy operator T.

Definition 11.1 — A **filtered** (φ, N)**-module** over K is the data of a filtered φ-module D over K together with a K_0-linear endomorphism $N: D \to D$, subject to the compatibility condition $N\varphi = p\varphi N$. $\diamond$

The notion of morphisms between filtered (φ, N)-modules is then the evident one: linear maps of K_0-vector spaces that commute with the respective φ, N, $F^\bullet$. Just as for filtered φ-modules, we do not obtain an abelian category: it's again the filtration's fault. We adjust for this using slopes, just as in section 8B.3. We thus say a filtered (φ, N)-module D is **weakly admissible** if it is μ-semistable of slope 0, with $\mu = \mu_{\mathrm{H}} - \mu_{\mathrm{N}}$ the interpolated slope function. In other words, this is again the requirement that $t_{\mathrm{N}}(D) = t_{\mathrm{H}}(D)$, and that for all subobjects C of D, $t_{\mathrm{N}}(D) \geqslant t_{\mathrm{H}}(D)$. Note however that this is weaker than asking that the underlying filtered φ-module be weakly admissible, because the condition applies only to subobjects in the category of filtered (φ, N)-modules, i.e. only to N-stable subspaces.

Again, by the general formalism of slope filtrations (in particular Proposition 8B.9), the category of weakly admissible filtered (φ, N)-modules is abelian. We still need Totaro's result [Totaro. $D \otimes D'$, Théorem 1] to show that this category is in fact Tannakian, as it is not evident that the tensor product of μ-semistable objects is still μ-semistable.

As with the crystalline comparison theorem, the semistable comparison theorem (Theorem 10.1) allows us to go between the $\mathbb{Q}_p[\mathrm{G}_K]$-module $V = \mathrm{H}^i_{\text{ét}}\left(X_{\bar{K}}, \mathbb{Q}_p\right)$ and the filtered (φ, N)-module over K given by $D = \mathrm{H}^i_{\mathrm{cris}}\left(Y^\wedge/W^\circ, \mathcal{O}_{Y^\wedge/W^\circ}\right) \otimes_W K_0$, for X/K a smooth proper scheme with semistable integral model $\mathcal{X}$ with special fibre Y.
Indeed, $D \cong \left(V \otimes_{\mathbb{Q}_p} \mathbf{B}_{\mathrm{st}}\right)^{\mathrm{G}_K}$ using C_{st} because $(\mathbf{B}_{\mathrm{st}})^{\mathrm{G}_K} = K_0$, and $V \cong F^0\left(D \otimes_{K_0} \mathbf{B}_{\mathrm{st}}\right)^{\varphi=1, N=0}$ using $\mathrm{C}_{\mathrm{st}}^{-1}$ as $F^0(\mathbf{B}_{\mathrm{st}})^{\varphi=1,N=0} = \mathbb{Q}_p$, as we saw at the end of section 9.10.

12 The main theorem of p-adic Hodge theory

We now expect Fontaine's theorem in the crystalline case (Theorem 8C.1) to generalise to the semistable case. This is indeed the case!

Théorem 12.1 (Colmez, Fontaine) — The functors D_{st} and V_{st} are quasi-inverse equivalences of Tannakian categories

$$\left\{\begin{array}{c}\text{Finite-dimensional continuous} \\ \text{semistable } \mathbb{Q}_p\text{-representations of } \mathrm{G}_K\end{array}\right\} \underset{\mathrm{V}_{\mathrm{cris}}}{\overset{\mathrm{D}_{\mathrm{cris}}}{\rightleftarrows}} \left\{\begin{array}{c}\text{Weakly admissible filtered} \\ (\varphi, N)\text{-modules over } K_0\end{array}\right\}. \quad \diamond$$

This theorem follows from a more precise statement:

Théorem 12.2 (Colmez, Fontaine) — Let D be a filtered (φ, N)-module of dimension $d \geqslant 1$. Then:

- $\dim_{\mathbb{Q}_p} \mathrm{V}_{\mathrm{st}}(D) < \infty$ if and only if $t_{\mathrm{N}}(C) \geqslant t_{\mathrm{H}}(C)$ for all subobjects $C \subseteq D$.
- Moreover, $\dim_{\mathbb{Q}_p} \mathrm{V}_{\mathrm{st}}(D) = d$ if and only if D is weakly admissible, i.e. in addition we also have $t_{\mathrm{N}}(D) = t_{\mathrm{H}}(D)$. In this case, D is admissible, which is to say it is in the essential image of the functor D_{st}. $\diamond$

Proof: This is [Colmez–Fontaine, Théorème 4.3]. See [Colmez–Fontaine, section 6]; the proof proceeds by the construction, for each weakly admissible filtered (φ, N)-module, of a different filtered (φ, N)-module by modifying the filtration. Then they prove that this second object is admissible [Colmez–Fontaine, Corollaire 6.3], which in turn can be shown to imply admissibility of the original filtered (φ, N)-module [Colmez–Fontaine, Proposition 6.1]. □

Note that this theorem provides an alternative proof that the category of weakly admissible filtered (φ, N)-modules is Tannakian. The difficult part is to check that the tensor product of two weakly admissible filtered (φ, N)-modules is itself weakly admissible; here this can be observed by marshalling the data across the functors D_{st} and V_{st}, which we know are compatible with tensor products.

13 The p-adic monodromy theorem

The crystalline comparison theorem $\mathrm{C}_{\mathrm{cris}}$ (Theorem 6.1) guarantees that the Galois representation afforded by the p-adic étale cohomology of a variety with good reduction is crystalline. We then extended this context to that of semistable schemes, with the introduction of $\mathbf{B}_{\mathrm{st}}$ and the notion of semistable p-adic Galois representations. This resulted in the statement of the semistable comparison theorem C_{st} (Theorem 10.1).
Now, in both the complex and ℓ-adic situations, this is not far from the general case. In the complex case, we saw in section 9.5.2 that the monodromy automorphism on the cohomology of a general family around a singular fibre is quasi-unipotent. In the ℓ-adic case, Grothendieck's ℓ-adic monodromy theorem (Theorem 9.1) guaranteed essentially the same thing: all ℓ-adic Galois representations of G_K are quasi-unipotent. We are thus led to believe that the same might hold in the p-adic case. This however is not true for an arbitrary p-adic representation; just as in the complex case we can't work with arbitrary variations of Hodge structure (with polarisability of the variation being the apposite algebraicity condition to impose), in the world of p-adic Hodge theory it is the condition that the representation be de Rham that is most relevant.

Theorem 13.1 (Berger, André, Kedlaya, Mebkhout) — All de Rham p-adic Galois representations of G_K are potentially semistable. ◇□

14 Semistability versus weights

Given a filtered (φ, N)-module D, we are interested in studying the interaction of N with the other structures. In particular, consider the problem of fixing the slopes of the Hodge polygon (the Hodge–Tate weights) and studying the allowable properties of N. The fundamental consideration is the following elementary proposition:

Proposition 14.1 — Let D be a (φ, N)-module over K_0. If $G_{\leqslant \bullet}$ denotes the ascending Dieudonné–Manin slope filtration (Theorem 8B.6) of the underlying φ-module, we have that $NG_{\leqslant\alpha} D \subseteq G_{\leqslant \alpha-1} D$. ◇

Proof: To simplify the situation, we might as well base change from K_0 to $(K_0^{\mathrm{nr}})^\vee$. Once this is done, the Dieudonné–Manin filtration is particularly easy to understand, as it exhibits every φ-module as a direct sum of simple μ-stable objects $\mathrm{D}_{r,s} = (K_0^{\mathrm{nr}})^\vee[\varphi]/(\varphi^r - p^s)$ with $\gcd(r,s) = 1$, and this is of slope $\alpha = {}^s/_r$. It then suffices to understand how N interacts with such a decomposition.
Let us begin by picking a vector in $v \in D_{(K_0^{\mathrm{nr}})^\vee}$ such that $v, \varphi v, \ldots, \varphi^{r-1} v$ span a copy of $\mathrm{D}_{r,s}$. If $Nv = 0$, there is nothing to show, so we may as well assume otherwise. From the relation $N\varphi = p\varphi N$, we deduce

$$\varphi^r N v = p^{-r} N \varphi^r v = p^{s-r} N v.$$

Therefore $Nv, \varphi Nv, \ldots, \varphi^{r-1} Nv$ span a copy of $\mathrm{D}_{r,s-r}$ inside $D_{(K_0^{\mathrm{nr}})^\vee}$. As ${}^{s-r}/_r = \alpha - 1$, we have that $Nv \in G_{\leqslant \alpha-1} D_{(K_0^{\mathrm{nr}})^\vee}$, and the result follows. □

We deduce in particular that N is necessarily nilpotent, as the Dieudonné–Manin filtration is finite. It also shows that N maps $\mathrm{gr}^G_\alpha D \to \mathrm{gr}^G_{\alpha-1} D$.

What do we gain from this? The idea is that N being non-zero forces there to be a pair of slopes $\alpha - 1, \alpha$. More generally, for every Jordan block of N of rank r, there will be a chain of slopes $\beta, \beta + 1, \ldots, \beta + r$. We can then leverage this if the Hodge polygon is fixed, as long chains of slopes might preclude the endpoints of the Newton and Hodge polygons from coinciding.
We start with a lemma, giving us a Jordan normal form for endomorphisms of a graded vector space.

Lemma 14.2 — Let $V = \bigoplus_i V_i$ be a graded vector space, and $N \circlearrowright V$ a homogeneous nilpotent endomorphism. There is then a basis of V consisting of homogeneous generalised eigenvectors for N. ◇

Proof: For the case without the grading, one way to proceed is to use the monodromy filtration $M_i V = \sum_{j-k=i} \ker(N^{j+1}) \cap \operatorname{im}(N^k)$. A splitting of this filtration then yields the requisite Jordan chains for N. For the same procedure to apply with the grading, we need to know this filtration $M_\bullet$ is compatible with the grading, meaning that $M_j V = \bigoplus_i M_j V_i$. This follows from homogeneity of N: N maps V_i to V_{i+d} for some d, so that $\ker(N^j) = \bigoplus_i \ker(N^j|_{V_i})$ and $\operatorname{im}(N^k) = \bigoplus_i \operatorname{im}(N^k|_{V_i})$. Therefore $\ker(N^{j+1}) \cap \operatorname{im}(N^k)$, and thus the whole filtration $M_\bullet$, are compatible with the grading. This ensures we can split the filtration $M_\bullet$ using homogeneous vectors, giving the required Jordan basis for N. □

Definition 14.3 — Let N be a nilpotent endomorphism on a vector space V. The **nilpotent rank** $\operatorname{nrk}(N)$ of N is the sum $\operatorname{nrk}(N) = \sum_{i \geqslant 1} \operatorname{rk}(N^i)$. Equivalently, $\operatorname{nrk}(N) = \sum_i \frac{r_i(r_i+1)}{2}$, where i indexes the Jordan blocks $\{J_i\}_i$ of N, with J_i of rank r_i. ◇

Theorem 14.4 — Let D be a weakly-admissible filtered (φ, N)-module over K, of dimension n. Let $h_1 \leqslant \cdots \leqslant h_n$ be the Hodge–Tate weights of D. Then $\operatorname{nrk}(N) \leqslant \sum_{i=1}^n (h_i - h_1)$. ◇

Proof: Consider first the underlying (φ, N)-module over K_0 of D, momentarily forgetting the Hodge filtration. Let $G_{\leqslant \bullet}$ be the (ascending) Dieudonné–Manin slope filtration attached to the Newton slope function given by Theorem 8B.6, and let $V = \operatorname{gr}^G D = \bigoplus_\alpha V_\alpha$. Proposition 14.1 guarantees that N acts on V, and is homogeneous of degree -1, i.e. $N\colon V_\alpha \to V_{\alpha-1}$.
Now we use the previous lemma to obtain a basis of homogeneous generalised eigenvectors for the action of N on V. We thus obtain Jordan chains $v_{i,0}, \ldots, v_{i,r_i}$ with $v_{i,k} \in \operatorname{gr}^G_{\alpha_i + r_i - k} D$ (so that v_{i,r_i} has slope α_i, and then v_{i,r_i-1} has slope $\alpha_i + 1$, and so on, up to $v_{i,0}$ of slope $\alpha_i + r_i$). Ordering the Jordan chains such that $\alpha_1 \leqslant \alpha_2 \leqslant \cdots \leqslant \alpha_m$, this shows:

$$\mathrm{t_N}(D) = \sum_\alpha \alpha \cdot \dim \operatorname{gr}^G_\alpha D = \sum_{i=1}^m \sum_{k=0}^{r_i} \alpha_i + k \geqslant \sum_{i=1}^m \sum_{k=0}^{r_i} \alpha_1 + k = n\alpha_1 + \operatorname{nrk}(N).$$

We now bring back in the Hodge filtration. As D is weakly admissible, we know that $\mathrm{t_N}(D) = \mathrm{t_H}(D)$; moreover $\alpha_1 \geqslant h_1$. We thus obtain:

$$\mathrm{t_H}(D) = \sum_{i=1}^n h_i \geqslant n h_1 + \operatorname{nrk}(N).$$ □

We can apply this result in a few examples.

- Modular forms.
  The local Galois representations associated with modular eigenforms of weight $k \geqslant 1$ have Hodge–Tate weights 0 and $k-1$. These Galois representations are de Rham; in the case that they are in fact semistable (which can always be arranged by taking a finite extension), we thus end up considering (weakly admissible) 2-dimensional filtered (φ, N)-modules over $\mathbb{Q}_p$, with Hodge–Tate weights 0 and $k-1$.
  Applying Theorem 14.4, it follows that if $k = 1$, then $\operatorname{nrk}(N) = 0$ and thus $N = 0$.
  As the Hodge–Tate weights aren't affected by finite base-change, this result implies that if $k = 1$, these Galois representations are potentially crystalline.

- Hilbert modular forms.
  Associated to a Hilbert modular eigenform f of weight $((k_\tau)_\tau, w)$ (with $k_\tau \geqslant 1$ and $k_\tau \equiv w \pmod 2$)

over the totally real field F, one obtains p-adic Galois representations of the form $\rho_{f,E}: \mathrm{G}_F \to \mathrm{GL}_2(E)$, for $E/\mathbb{Q}_p$ some finite extension [Jarvis. $\rho_{f,\ell}$, Theorem 6.1]. Restricting to the decomposition subgroup at some prime $\mathfrak{p}$ above p, we obtain a local Galois representation $\rho_{f,\mathfrak{p},E}: \mathrm{G}_K \to \mathrm{GL}_2(E)$, where $K = F_{\mathfrak{p}}$ is the completion of F at $\mathfrak{p}$; we assume that E contains the Galois closure of K.

If the weight of f is **regular**, meaning that $k_\tau \geqslant 2$ for all τ, then $\rho_{f,\mathfrak{p},E}$ is known to be potentially semistable, with Hodge–Tate weights as below [Skinner. Note, Theorem 1]. However, this is only conjectural in general. Assuming that $\rho_{f,\mathfrak{p},E}$ is semistable, application of D_{st} then yields a (weakly admissible) filtered (φ, N)-module over K with additional action of E, i.e. a (φ, N)-module over $K_0 \otimes_{\mathbb{Q}_p} E$, free of rank 2, with a filtration defined after tensoring to K (and satisfying weak admissibility). The labelled Hodge–Tate weights of this object are expected to be (and known to be, if the weight is regular)

$$\left(\frac{w-k_i}{2}, \frac{w+k_i}{2}-1\right)_i,$$

with i indexing the embeddings of K into E. The $(k_i)_{\tau_i: K \hookrightarrow E}$ that appear here are a subset of the $(k_\tau)_{\tau: F \hookrightarrow \overline{\mathbb{Q}}}$, the subset being determined by the choice of $\mathfrak{p}$, and w is as above.

Theorem 14.4 then yields

$$[E:K] \sum_{i=1}^{[K:\mathbb{Q}_p]} (w - 1 - (w - k_{\max})) \geqslant \mathrm{nrk}(N).$$

As N is nilpotent and E-linear, either it is 0 or $\mathrm{nrk}(N) = [E : \mathbb{Q}_p]$. Thus, a necessary condition for $N \neq 0$ is

$$[E:K][K:\mathbb{Q}_p](k_{\max} - 1) \geqslant [E:\mathbb{Q}_p],$$

which is simply the condition $k_{\max} \geqslant 2$.

We can, however, get a better result than this. Twisting $D_{f,\mathfrak{p}}$ by Lubin–Tate characters, we can assume the labelled Hodge–Tate weights are $(0, k_i - 1)_i$. Applying Theorem 14.4 in this case, we arrive at the inequality

$$[E:K] \sum_{i=1}^{[K:\mathbb{Q}_p]} (k_i - 1) \geqslant [E:\mathbb{Q}_p].$$

As $D_{f,\mathfrak{p}}$ supports a nontrivial N if and only if its twists do, this means that a necessary condition for $N \neq 0$ is that $\sum_i k_i \geqslant 2[K:\mathbb{Q}_p]$, i.e. that the k_i average at least 2.

Again, as this result depends only on the Hodge–Tate weights, this implies that if the k_i do not average at least 2, $\rho_{f,\mathfrak{p},E}$ is potentially crystalline (if it is indeed de Rham with the expected Hodge–Tate weights).

15 p-adic Hodge theory in rank 1

It will be useful to record a few observations about rank 1 filtered φ-modules.

Naturally, this means we are considering p-adic Galois characters $\psi: \mathrm{G}_K \to E^\times$, for $E/\mathbb{Q}_p$ some finite extension. To fit these into the framework of p-adic Hodge theory, the natural questions to ask are: which of these characters are de Rham? Which are crystalline?

The classification of characters of G_K proceeds via local class field theory. The local reciprocity law hinges on the local Artin map $\mathrm{rec}_K: K^\times \to \mathrm{G}_K^{\mathrm{ab}}$, which we normalise arithmetically: rec_K maps ϖ to the image of the arithmetic Frobenius in $\mathrm{G}_K^{\mathrm{ab}}$, denoted Frob_q. Using rec_K, we can then give the following criterion for admissibility of p-adic characters.

Proposition 15.1 — Let $\psi: \mathrm{G}_K \to E^\times$ be a continuous character. Then:

- ψ is de Rham if and only if $\psi \circ \mathrm{rec}_K: K^\times \to E^\times$ is locally algebraic,
- ψ is crystalline if and only if $(\psi \circ \mathrm{rec}_K)|_{\mathcal{O}_K^\times}: \mathcal{O}_K^\times \to E^\times$ is algebraic. ◇

Proof: See [Conrad. Local χ, Proposition B.4, (i)]. □

The algebraicity conditions are to be considered with respect to the corresponding algebraic groups over $\mathbb{Q}_p$, $\mathrm{Res}^K_{\mathbb{Q}_p}\mathbb{G}_{m,K}$ and $\mathrm{Res}^E_{\mathbb{Q}_p}\mathbb{G}_{m,E}$, whose sets of $\mathbb{Q}_p$-points are $K^\times$ and $E^\times$, respectively. Thus $\zeta\colon K^\times \to E^\times$ is locally algebraic if $\zeta|_{\varpi^n \mathcal{O}_K^\times} = \xi|_{\varpi^n \mathcal{O}_K^\times}$ for some integer $n \geqslant 0$ and some algebraic $\xi\colon \mathrm{Res}^K_{\mathbb{Q}_p}\mathbb{G}_{m,K} \to \mathrm{Res}^E_{\mathbb{Q}_p}\mathbb{G}_{m,E}$.
In addition to the above result, when ψ is crystalline, we can explicitly describe $\mathrm{D}_{\mathrm{cris}}(E(\psi))$.

Proposition 15.2 — Assume that $\psi\colon \mathrm{G}_K \to E^\times$ is crystalline, so that $(\psi \circ \mathrm{rec}_K)|_{\mathcal{O}_K^\times} = \xi|_{\mathcal{O}_K^\times}$ for some algebraic character $\xi\colon K^\times \to E^\times$.
The filtered φ-module $\mathrm{D}_{\mathrm{cris}}(E(\psi))$ is a free rank 1 module over $K_0 \otimes_{\mathbb{Q}_p} E$ with

$$\varphi^a = 1 \otimes \Big(\psi(\mathrm{rec}_K(\varpi))^{-1} \cdot \xi(\varpi)\Big),$$

where $a = [K_0 : \mathbb{Q}_p]$. Enlarging E if necessary, so that it contains the Galois closure of K, we can write $\xi = \prod_\tau \tau^{n_\tau}$ with τ running over the embeddings $K \hookrightarrow E$. The labelled Hodge–Tate weights of $\mathrm{D}_{\mathrm{cris}}(E(\psi))$ are then $(n_\tau)_\tau$. ◇

Proof: Following [Conrad. Local χ, Proposition B.4, (ii)], the plan is to reduce to the case where all Hodge–Tate weights, except one, vanish. To tackle this case, the idea is to use a certain free $\mathcal{O}_K$-module of rank 1 on which G_K acts: the p-adic Tate module of a formal $\mathcal{O}_K$-module over K, the **Lubin–Tate formal group** $\Gamma_{K,\varpi}$ attached to the pair (K, ϖ). This is the (unique up to isomorphism) 1-dimensional p-divisible formal $\mathcal{O}_K$-module over $\mathcal{O}_K$ such that multiplication by ϖ reduces modulo ϖ to the Frobenius automorphism Frob_q [Lubin–Tate, Theorem 1]. Write $\mathrm{H}_{\kappa,\varpi}$ for the reduction of $\Gamma_{K,\varpi}$, the **Honda formal group**. It is the (unique up to isomorphism) 1-dimensional p-divisible formal $\mathcal{O}_K$-module over κ with $[\varpi](T) = T^q$ (and is thus of height $e \cdot a$, as then $[p](T) = T^{q^e} = T^{p^{ae}}$).[62]
The Galois action on the Tate module of $\Gamma_{K,\varpi}$ thus yields a crystalline character $\chi_{K,\varpi}\colon \mathrm{G}_K \to \mathcal{O}_K^\times$. A fundamental calculation then describes $\Big(\chi_{K,\varpi} \circ \mathrm{rec}_K\Big)\Big|_{\mathcal{O}_K^\times}\colon \mathcal{O}_K^\times \to \mathcal{O}_K^\times$ as given by inversion $(-)^{-1}\colon \mathcal{O}_K^\times \to \mathcal{O}_K^\times$; also $\chi_{K,\varpi}(\mathrm{Frob}_\varpi) = 1$ [Lubin–Tate, Theorem 2]. Post-composing with an embedding $\tau\colon K \hookrightarrow E$, $\tau \circ \chi_{K,\varpi}$ is then the unique crystalline character $\mathrm{G}_K \to \mathcal{O}_K^\times$ which is trivial on Frob_ϖ and with Hodge–Tate weight -1 at the chosen embedding τ (and other Hodge–Tate weights 0). The sign is due to the inversion above.
On the other hand, we already mentioned that multiplication by ϖ induces, on the reduction $\mathrm{H}_{\kappa,\varpi}$ of $\Gamma_{K,\varpi}$, the Frobenius automorphism Frob_q. This allows us to deduce, using the compatibility between $\mathrm{D}_{\mathrm{cris}}$ and Dieudonné modules [Conrad. Local χ, Example B.5], the action of Frobenius on $\mathrm{D}_{\mathrm{cris}}(\chi_\varpi)$. We have an isomorphism of φ-modules over $K_0 \otimes_{\mathbb{Q}_p} K$

$$(\mathrm{D}^*(\mathrm{H}_{\kappa,\varpi}))^\vee \cong \mathrm{D}_{\mathrm{cris}}(\chi_{K,\varpi}),$$

with $\mathrm{D}^*(\mathrm{H}_{\kappa,\varpi})$ denoting the contravariant Dieudonné module of $\mathrm{H}_{\kappa,\varpi}$. Write F for the Frobenius on $\mathrm{D}^*(\mathrm{H}_{\kappa,\varpi})$ and φ for the Frobenius on $\mathrm{D}_{\mathrm{cris}}(\chi_\varpi)$. From the above, we obtain that $F^a = 1 \otimes \varpi$. On the dual φ-module, φ acts by the formula $\varphi(f) = \sigma \circ f \circ F^{-1}$, so that φ^a acts by $1 \otimes \varpi^{-1}$. This is compatible with the formula $\varphi^a = 1 \otimes \Big(\psi(\mathrm{rec}_K(\varpi))^{-1} \cdot \xi(\varpi)\Big)$ of the proposition, as in this case $\psi(\mathrm{rec}_K(\varpi)) = \chi_{K,\varpi}(\mathrm{Frob}_\varpi) = 1$ and $\xi = (-)^{-1}$.
The general case is finally tackled by taking products of Lubin–Tate characters: every crystalline character $\psi\colon \mathrm{G}_K \to E$ can be uniquely expressed as a product of Lubin–Tate characters, times an unramified character. This statement can be deduced from the above: twisting by Lubin–Tate characters, we can arrange that all the Hodge–Tate weights are 0, obtaining an unramified character.
Using multiplicativity of the proposed formula for φ^a, and tensor compatibility of $\mathrm{D}_{\mathrm{cris}}$, it thus only remains to show that the formula holds in the case that ψ is unramified. In this case, $\xi = 1$ and the formula reads $\varphi^a = 1 \otimes \psi(\mathrm{rec}_K(\varpi))^{-1}$. This formula can be verified by using the functor D_{nr} associated with the period ring $\mathbf{B}_{\mathrm{nr}} = (K_0^{\mathrm{nr}})^\vee \subset \mathbf{B}_{\mathrm{cris}}$; the sign coming once again from the fact that we are using a covariant and not contravariant functor. □

Remark 15.3 – In this situation, the knowledge of φ^a uniquely determines the isomorphism class of the corresponding φ-module. To see this, suppose we are given two injective σ-semilinear endomorphisms of

[62] Recall that $a = [K_0 : \mathbb{Q}_p]$, $e = [K : K_0]$, $q = p^a$, $(p) = (\varpi^e)$.

a free $K_0 \otimes_{\mathbb{Q}_p} E$-module of rank 1, say φ_1 and φ_2, with $\varphi_1^a = \varphi_2^a$. Choosing a basis vector e, we write $\varphi_1 e = \lambda_1 e$ and $\varphi_2 e = \lambda_2 e$ for some $\lambda_1, \lambda_2 \in (K_0 \otimes_{\mathbb{Q}_p} E)^\times$. The condition that $\varphi_1^a = \varphi_2^a$ implies that $\lambda_1 \sigma(\lambda_1) \cdots \sigma^{a-1}(\lambda_1) = \lambda_2 \sigma(\lambda_2) \cdots \sigma^{a-1}(\lambda_2)$, i.e. λ_1 and λ_2 have equal norm. Hilbert's theorem 90 (applied to the algebraic group $\mathrm{Res}^E_{\mathbb{Q}_p} \mathbb{G}_{\mathrm{m},E}$) then guarantees that there exists $\gamma \in (K_0 \otimes_{\mathbb{Q}_p} E)^\times$ such that $\frac{\sigma(\gamma)}{\gamma} = \frac{\lambda_1}{\lambda_2}$. Using the basis (γe), we see that $\varphi_2(\gamma e) = \sigma(\gamma)\varphi_2 e = \frac{\sigma(\gamma)}{\gamma}\lambda_2(\gamma e) = \frac{\lambda_1}{\lambda_2}\lambda_2(\gamma e) = \lambda_1(\gamma e)$. The change of basis $e \rightsquigarrow \gamma e$ thus shows that φ_1 and φ_2 determine isomorphic φ-modules.

16 Bloch–Kato theory

In section 5.5, we described through some exact sequences the interaction of the Frobenius endomorphism, the Galois action, and the Hodge filtration on $\mathbf{B}_{\mathrm{cris}}$, which we extended in section 9.10 to $\mathbf{B}_{\mathrm{st}}$. To recapitulate, we have a double complex and a triple complex whose total complexes are exact. One corresponding to $\mathbf{B}_{\mathrm{cris}}$:

$$\begin{tikzcd}
\mathbb{Q}_p \arrow[r] \arrow[d] & \mathbf{B}_{\mathrm{cris}} \arrow[r, "\varphi-1"] \arrow[d] & \mathbf{B}_{\mathrm{cris}} \\
\mathbf{B}^+_{\mathrm{dR}} \arrow[r] & \mathbf{B}_{\mathrm{dR}} &
\end{tikzcd}$$

and one corresponding to $\mathbf{B}_{\mathrm{st}}$:

$$\begin{tikzcd}
\mathbb{Q}_p \arrow[rr] \arrow[ddd] & & \mathbf{B}_{\mathrm{st}} \arrow[rr, "\varphi-1"] \arrow[dl, "N"'] \arrow[ddd] & & \mathbf{B}_{\mathrm{st}} \arrow[dl, "N"] \\
& \mathbf{B}_{\mathrm{st}} \arrow[rr, "1-p\varphi"] & & \mathbf{B}_{\mathrm{st}} & \\
& & & & \\
\mathbf{B}^+_{\mathrm{dR}} \arrow[rr] & & \mathbf{B}_{\mathrm{dR}} & &
\end{tikzcd}$$

This means that, in a derived sense, we can consider that $\mathbb{Q}_p$ bears a family resemblance with the (double/triple) complexes

$$\begin{tikzcd}
& \mathbf{B}_{\mathrm{cris}} \arrow[r, "\varphi-1"] \arrow[d] & \mathbf{B}_{\mathrm{cris}} \\
\mathbf{B}^+_{\mathrm{dR}} \arrow[r] & \mathbf{B}_{\mathrm{dR}} &
\end{tikzcd}$$

and

$$\begin{tikzcd}
& & \mathbf{B}_{\mathrm{st}} \arrow[rr, "\varphi-1"] \arrow[dl, "N"'] \arrow[ddd] & & \mathbf{B}_{\mathrm{st}} \arrow[dl, "N"] \\
& \mathbf{B}_{\mathrm{st}} \arrow[rr, "1-p\varphi"] & & \mathbf{B}_{\mathrm{st}} & \\
& & & & \\
\mathbf{B}^+_{\mathrm{dR}} \arrow[rr] & & \mathbf{B}_{\mathrm{dR}} & &
\end{tikzcd}$$

or their total complexes.

This is very useful from the point of view of p-adic Galois representations. It means that instead of considering the Galois cohomology of a crystalline $\mathbb{Q}_p$-representation V of G_K, we could instead consider the cohomology of the double complex

$$\begin{tikzcd}
& \mathrm{D}_{\mathrm{cris}}(V) \arrow[r, "\varphi-1"] \arrow[d] & \mathrm{D}_{\mathrm{cris}}(V) \\
F^0\mathrm{D}_{\mathrm{dR}}(V) \arrow[r] & \mathrm{D}_{\mathrm{dR}}(V) &
\end{tikzcd}$$

This leads us to define functors $\mathcal{K}^\bullet_{\mathrm{cris}}$ and $\mathcal{K}^\bullet_{\mathrm{st}}$. Given a filtered φ-module D, $\mathcal{K}^\bullet_{\mathrm{cris}}(D)$ is the total complex of the double complex

$$\begin{tikzcd} & D \arrow[r, "\varphi-1"] \arrow[d] & D \\ F^0 D_K \arrow[r] & D_K & \end{tikzcd}$$

Similarly, if D is instead a filtered (φ, N)-module, $\mathcal{K}^\bullet_{\mathrm{st}}(D)$ is the total complex of the triple complex

$$\begin{tikzcd} & & D \arrow[rr, "\varphi-1"] \arrow[dl, "N"'] \arrow[d, dash] & & D \arrow[dl, "N"] \\ & D \arrow[rr, "1-p\varphi"] & \arrow[d] & D & \\ F^0 D_K \arrow[rr] & & D_K & & \end{tikzcd}$$

These are complexes of $\mathbb{Q}_p$-vector spaces.
In both of these cases, we are placing the initial D and $F^0 D_K$ in degree 0.[63] Concretely:

$$\mathcal{K}^\bullet_{\mathrm{cris}}(D) = \Big[\overset{0}{D \oplus F^0 D_K} \xrightarrow{i} \overset{1}{D \oplus D_K} \Big],$$

with the map i given by $i(d_1, d_2) = (\varphi d_1 - d_1, d_1 - d_2)$, where d_1, d_2 are considered as elements of D_K through the natural inclusions $\mathbf{B}_{\mathrm{cris}} \otimes_{K_0} K \hookrightarrow \mathbf{B}_{\mathrm{dR}}$, and $\mathbf{B}^+_{\mathrm{dR}} \hookrightarrow \mathbf{B}_{\mathrm{dR}}$.
We can also describe $\mathcal{K}^\bullet_{\mathrm{st}}$ in this way:

$$\mathcal{K}^\bullet_{\mathrm{st}}(D) = \Big[\overset{0}{D \oplus F^0 D_K} \xrightarrow{j} \overset{1}{D \oplus D \oplus D_K} \xrightarrow{k} \overset{2}{D} \Big],$$

with $j(d_1, d_2) = (\varphi d_1 - d_1, N d_1, d_1 - d_2)$, $k(e_1, e_2, e_3) = N e_1 + e_2 - p\varphi e_2$. Here we need to be careful to pick an inclusion $\mathbf{B}_{\mathrm{st}} \to \mathbf{B}_{\mathrm{dR}}$, by choosing a value for $\log(p)$ as we saw in section 9.10.

The upshot of all of this, as we might expect from the previous intuitive observations, is that we should be able to compute the Galois cohomology of a p-adic Galois representation V through the cohomology of $\mathcal{K}^\bullet_{\mathrm{cris}}(\mathrm{D}_{\mathrm{cris}}(V))$ if V is crystalline, or of $\mathcal{K}^\bullet_{\mathrm{st}}(\mathrm{D}_{\mathrm{st}}(V))$ if V is semistable. As an immediate corollary of Corollary 5.7, we have:

Corollary 16.1 — The complexes $\mathcal{K}^\bullet_{\mathrm{cris}}(\mathbf{B}_{\mathrm{cris}})$ and $\mathcal{K}^\bullet_{\mathrm{st}}(\mathbf{B}_{\mathrm{st}})$ are both quasi isomorphic to $\mathbb{Q}_p$, concentrated in degree 0. ◇□

We can now make use of the functors $\mathcal{K}_{\mathrm{cris}}$ and $\mathcal{K}_{\mathrm{st}}$ to compute Galois cohomology of Galois representations. Given a Galois representation $\rho\colon \mathrm{G}_K \to V$, we define:

$$\begin{aligned}
\mathrm{H}^1_{\mathrm{nr}}(\mathrm{G}_K, V) &= \ker\left(\mathrm{H}^1(\mathrm{G}_K, V) \longrightarrow \mathrm{H}^1\left(\mathrm{G}_K, V \otimes_{\mathbb{Q}_p} (K_0^{\mathrm{nr}})^\vee\right)\right), \\
\mathrm{H}^1_{\mathrm{e}}(\mathrm{G}_K, V) &= \ker\left(\mathrm{H}^1(\mathrm{G}_K, V) \longrightarrow \mathrm{H}^1\left(\mathrm{G}_K, V \otimes_{\mathbb{Q}_p} \mathbf{B}^{\varphi=1}_{\mathrm{cris}}\right)\right), \\
\mathrm{H}^1_{\mathrm{f}}(\mathrm{G}_K, V) &= \ker\left(\mathrm{H}^1(\mathrm{G}_K, V) \longrightarrow \mathrm{H}^1\left(\mathrm{G}_K, V \otimes_{\mathbb{Q}_p} \mathbf{B}_{\mathrm{cris}}\right)\right), \\
\mathrm{H}^1_{\mathrm{g}}(\mathrm{G}_K, V) &= \ker\left(\mathrm{H}^1(\mathrm{G}_K, V) \longrightarrow \mathrm{H}^1\left(\mathrm{G}_K, V \otimes_{\mathbb{Q}_p} \mathbf{B}_{\mathrm{dR}}\right)\right), \\
\mathrm{H}^1_{\mathrm{h}}(\mathrm{G}_K, V) &= \ker\left(\mathrm{H}^1(\mathrm{G}_K, V) \longrightarrow \mathrm{H}^1\left(\mathrm{G}_K, V \otimes_{\mathbb{Q}_p} \mathbf{B}^{\varphi=1}_{\mathrm{st}}\right)\right), \\
\mathrm{H}^1_{\mathrm{st}}(\mathrm{G}_K, V) &= \ker\left(\mathrm{H}^1(\mathrm{G}_K, V) \longrightarrow \mathrm{H}^1\left(\mathrm{G}_K, V \otimes_{\mathbb{Q}_p} \mathbf{B}_{\mathrm{st}}\right)\right).
\end{aligned}$$

We can then compute the cohomology groups $\mathrm{H}^1_{\mathrm{f}}(\mathrm{G}_K, V)$, $\mathrm{H}^1_{\mathrm{st}}(\mathrm{G}_K, V)$ as the degree 1 cohomology of the complexes $\mathcal{K}_{\mathrm{cris}}(\mathrm{D}_{\mathrm{cris}}(V))$, $\mathcal{K}_{\mathrm{cris}}(\mathrm{D}_{\mathrm{st}}(V))$ [Fontaine–Perrin-Riou, 3.3.3, Remarque].
Variants are also possible for $\mathrm{H}^1_{\mathrm{nr}}(\mathrm{G}_K, V)$, $\mathrm{H}^1_{\mathrm{e}}(\mathrm{G}_K, V)$ and $\mathrm{H}^1_{\mathrm{h}}(\mathrm{G}_K, V)$ [Fontaine–Perrin-Riou, 3.3.2, 3.3.3], with their associated Fontaine–Dieudonné functors $\mathrm{D}_?$.

[63] This is similar to the process of replacing an object by a resolution, wherein a resolution

$$0 \to C \to I_0 \to I_1 \to I_2 \to \cdots$$

leads to the replacement of C (considered as a complex concentrated in degree 0) by the complex $[I_0 \to I_1 \to I_2 \to \cdots]$, with I_i in degree i.

This motivates us more generally to define the groups $\mathrm{H}^i_?(\mathrm{G}_K(V))$ as the degree i cohomology of the relevant complexes, when $i \neq 1$.
The utility of the groups $\mathrm{H}^1_?(\mathrm{G}_K, V)$ is that they control extensions. We have [Fontaine–Perrin-Riou, Proposition 3.3.7]

Proposition 16.2 — The subspaces $\mathrm{H}^1_?(\mathrm{G}_K, V) \subseteq \mathrm{H}^1(\mathrm{G}_K, V)$ classify the extensions

$$0 \to V \to W \to \mathbb{Q}_p \to 0$$

such that

$$0 \to \mathrm{D}_?(V) \to \mathrm{D}_?(W) \to \mathrm{D}_?(\mathbb{Q}_p) \to 0$$

is also exact, where $\mathrm{D}_?$ is the corresponding Fontaine–Dieudonné functor (e.g. $\mathrm{D}_\mathrm{f} = \mathrm{D}_\mathrm{cris}$).
In particular, if V is $\mathbf{B}_?$-admissible, then $\mathrm{H}^1_?(\mathrm{G}_K, V)$ classifies $\mathbf{B}_?$-admissible extensions of $\mathbb{Q}_p$ by V. ◇

16.1 Semistable non-crystalline extensions

Using the functors $\mathcal{K}^\bullet_\mathrm{cris}$ and $\mathcal{K}^\bullet_\mathrm{st}$ we can perform a hands-on calculation of the dimensions of the spaces H^i_f and H^i_st.
Suppose then that V is a crystalline representation of G_K, with $D = \mathrm{D}_\mathrm{cris}(V)$. From the exact sequence

$$\mathcal{K}^\bullet_\mathrm{cris}(D) = \Big[\overset{0}{D \oplus F^0 D_K} \xrightarrow{\ i\ } \overset{1}{D \oplus D_K} \Big],$$

with $i(d_1, d_2) = (\varphi d_1 - d_1, d_1 - d_2)$, we deduce

$$\begin{aligned} \mathrm{h}^0_\mathrm{f}(\mathrm{G}_K, V) &= \dim_{\mathbb{Q}_p} F^0 D^{\varphi=1}, \\ \mathrm{h}^1_\mathrm{f}(\mathrm{G}_K, V) &= \dim_{\mathbb{Q}_p} D_K / F^0 D_K + \dim_{\mathbb{Q}_p} F^0 D^{\varphi=1}. \end{aligned}$$

We thus obtain an expression for the Euler characteristic: $\chi_\mathrm{f}(\mathrm{G}_K, V) = -\dim_{\mathbb{Q}_p} D_K / F^0 D_K$.
One the other hand suppose V is semistable, and write $D = \mathrm{D}_\mathrm{st}(V)$. We then have:

$$\mathcal{K}^\bullet_\mathrm{st}(D) = \Big[\overset{0}{D \oplus F^0 D_K} \xrightarrow{\ j\ } \overset{1}{D \oplus D \oplus D_K} \xrightarrow{\ k\ } \overset{2}{D} \Big],$$

with $j(d_1, d_2) = (\varphi d_1 - d_1, N d_1, d_1 - d_2)$ and $k(e_1, e_2, e_3) = N e_1 + e_2 - p\varphi e_2$, from which follow

$$\begin{aligned} \mathrm{h}^0_\mathrm{st}(\mathrm{G}_K, V) &= \dim_{\mathbb{Q}_p} F^0 D^{\varphi=1, N=0}, \\ \mathrm{h}^1_\mathrm{st}(\mathrm{G}_K, V) &= \dim_{\mathbb{Q}_p} D_K / F^0 D_K + \dim_{\mathbb{Q}_p} \ker\left((1 - p\varphi) + N \colon D \oplus D \to D\right) \\ &\quad - \left(\dim_{\mathbb{Q}_p} D - \dim_{\mathbb{Q}_p} F^0 D^{\varphi=1, N=0}\right), \\ \mathrm{h}^2_\mathrm{st}(\mathrm{G}_K, V) &= \dim_{\mathbb{Q}_p} \ker\left((1 - p\varphi) + N \colon D \oplus D \to D\right) - \dim_{\mathbb{Q}_p} D. \end{aligned}$$

In this case, the Euler characteristic is $\chi_\mathrm{st}(\mathrm{G}_K, V) = -\dim_{\mathbb{Q}_p} D_K / F^0 D_K - \dim_{\mathbb{Q}_p} D$.
If V is moreover assumed to be crystalline, so that $N = 0$ in $\mathrm{D}_\mathrm{st}(V)$, we simplify:

$$\begin{aligned} \mathrm{h}^0_\mathrm{st}(\mathrm{G}_K, V) &= \dim_{\mathbb{Q}_p} F^0 D^{\varphi=1} = \mathrm{h}^0_\mathrm{f}(\mathrm{G}_K, V), \\ \mathrm{h}^1_\mathrm{st}(\mathrm{G}_K, V) &= \dim_{\mathbb{Q}_p} D_K / F^0 D_K + \dim_{\mathbb{Q}_p} D^{p\varphi=1} + \dim_{\mathbb{Q}_p} F^0 D^{\varphi=1}, \\ \mathrm{h}^2_\mathrm{st}(\mathrm{G}_K, V) &= \dim_{\mathbb{Q}_p} D^{p\varphi=1}. \end{aligned}$$

After rearranging we thus obtain:

$$\mathrm{h}^1_\mathrm{st}(\mathrm{G}_K, V) - \mathrm{h}^1_\mathrm{f}(\mathrm{G}_K, V) = \mathrm{h}^2_\mathrm{st}(\mathrm{G}_K, V).$$

This means that, by computing the dimension of $\mathrm{h}^2_\mathrm{st}(\mathrm{G}_K, V)$, we learn whether to expect semistable non-crystalline extensions of the trivial representation $\mathbf{1}$ by V.

We can refine this somewhat. Following [Nekovář. Heights, Theorem 1.17], consider again the double/triple complexes at the beginning of section 16. We could relate them by inscribing both into a large 4-dimensional diagram, but it is more helpful to summarise the two complexes by short exact sequences before relating them. For the crystalline double complex, we simply have:

$$0 \longrightarrow \mathbb{Q}_p \longrightarrow \mathbf{B}_{\mathrm{cris}} \oplus \mathbf{B}_{\mathrm{dR}}^+ \xrightarrow{(\varphi-1)\oplus\iota} \mathbf{B}_{\mathrm{cris}} \oplus \mathbf{B}_{\mathrm{dR}} \longrightarrow 0.$$

For the semistable triple complex, we let $\mathbf{B}_{\mathrm{st}}^0$ be the pullback:

$$\begin{tikzcd} \mathbf{B}_{\mathrm{st}}^0 \arrow[r, "\alpha"] \arrow[d, "\beta"'] & \mathbf{B}_{\mathrm{st}} \arrow[d, "N"] \\ \mathbf{B}_{\mathrm{st}} \arrow[r, "1-p\varphi"] & \mathbf{B}_{\mathrm{st}}. \end{tikzcd}$$

Exactness of the semistable triple complex thus yields the exactness of the short sequence

$$0 \longrightarrow \mathbb{Q}_p \longrightarrow \mathbf{B}_{\mathrm{st}} \oplus \mathbf{B}_{\mathrm{dR}}^+ \xrightarrow{\delta\oplus\iota} \mathbf{B}_{\mathrm{st}}^0 \oplus \mathbf{B}_{\mathrm{dR}} \longrightarrow 0,$$

where $\delta\colon \mathbf{B}_{\mathrm{st}} \to \mathbf{B}_{\mathrm{st}}^0$ is the unique homomorphism satisfying $\alpha \circ \delta = \varphi - 1$, $\beta \circ \delta = N$, obtained by the universal property of the pullback.
We can now put these two exact sequences together:

$$\begin{tikzcd} & & 0 \arrow[d] & 0 \arrow[d] & \\ 0 \arrow[r] & \mathbb{Q}_p \arrow[r] \arrow[d, equal] & \mathbf{B}_{\mathrm{cris}} \oplus \mathbf{B}_{\mathrm{dR}}^+ \arrow[r, "(\varphi-1)\oplus\iota"] \arrow[d] & \mathbf{B}_{\mathrm{cris}} \oplus \mathbf{B}_{\mathrm{dR}} \arrow[r] \arrow[d] & 0 \\ 0 \arrow[r] & \mathbb{Q}_p \arrow[r] & \mathbf{B}_{\mathrm{st}} \oplus \mathbf{B}_{\mathrm{dR}}^+ \arrow[r, "\delta\oplus\iota"] \arrow[d, "N"'] & \mathbf{B}_{\mathrm{st}}^0 \oplus \mathbf{B}_{\mathrm{dR}} \arrow[r] \arrow[d, "\beta"] & 0 \\ & & \mathbf{B}_{\mathrm{st}} \arrow[r, equal] \arrow[d] & \mathbf{B}_{\mathrm{st}} \arrow[d] & \\ & & 0 & 0 & \end{tikzcd}$$

The three rows and three columns of this diagram are already all known to be exact, except possibly the third column, which is then also guaranteed to be exact by the nine lemma. Moving over the factors of $\mathbf{B}_{\mathrm{dR}}^+$ to the third column (where we end up with summands $\mathbf{B}_{\mathrm{dR}}/\mathbf{B}_{\mathrm{dR}}^+$), tensoring with an arbitrary semistable representation V, and considering the associated long exact sequences in cohomology after applying the G_K-invariants functor, we then obtain the following commutative diagram with exact rows and columns (thanks to left-exactness of all functors involved):

$$\begin{tikzcd} & & 0 \arrow[d] & 0 \arrow[d] & & \\ 0 \arrow[r] & \mathrm{H}^0(\mathrm{G}_K, V) \arrow[r] \arrow[d, equal] & \mathrm{D}_{\mathrm{cris}}(V) \arrow[r, "(\varphi-1)\oplus\iota"] \arrow[d] & \mathrm{D}_{\mathrm{cris}}(V) \oplus \mathrm{D}_{\mathrm{dR}}(V)/F^0\mathrm{D}_{\mathrm{dR}}(V) \arrow[r] \arrow[d] & \mathrm{H}^1_{\mathrm{f}}(\mathrm{G}_K, V) \arrow[r] \arrow[d] & 0 \\ 0 \arrow[r] & \mathrm{H}^0(\mathrm{G}_K, V) \arrow[r] & \mathrm{D}_{\mathrm{st}}(V) \arrow[r, "\delta\oplus\iota"] \arrow[d, "N"'] & \mathrm{D}_{\mathrm{st}}^0(V) \oplus \mathrm{D}_{\mathrm{dR}}(V)/F^0\mathrm{D}_{\mathrm{dR}}(V) \arrow[r] \arrow[d, "\beta"] & \mathrm{H}^1_{\mathrm{st}}(\mathrm{G}_K, V) \arrow[r] & 0 \\ & & \mathrm{D}_{\mathrm{st}}(V) \arrow[r, equal] & \mathrm{D}_{\mathrm{st}}(V) & & \end{tikzcd}$$

Here $\mathrm{D}_{\mathrm{st}}^0(V) := \big(V \otimes_{\mathbb{Q}_p} \mathbf{B}_{\mathrm{st}}^0\big)^{\mathrm{G}_K}$. Doing a diagram chase along the diagram

$$\begin{tikzcd}
 & 0 \arrow[d] & & \\
 & \mathrm{D}_{\mathrm{cris}}(V) \oplus \mathrm{D}_{\mathrm{dR}}(V)/F^0\mathrm{D}_{\mathrm{dR}}(V) \arrow[r] \arrow[d] & \mathrm{H}^1_{\mathrm{f}}(\mathrm{G}_K, V) \arrow[r] \arrow[d] & 0 \\
\mathrm{D}_{\mathrm{st}}(V) \arrow[r, "\delta\oplus\iota"] \arrow[d, "N"'] & \mathrm{D}^0_{\mathrm{st}}(V) \oplus \mathrm{D}_{\mathrm{dR}}(V)/F^0\mathrm{D}_{\mathrm{dR}}(V) \arrow[r] \arrow[d, "\beta"] & \mathrm{H}^1_{\mathrm{st}}(\mathrm{G}_K, V) \arrow[r] & 0 \\
\mathrm{D}_{\mathrm{st}}(V) \arrow[r, equal] & \mathrm{D}_{\mathrm{st}}(V) & &
\end{tikzcd}$$

we obtain the short exact sequence

$$0 \longrightarrow \mathrm{H}^1_{\mathrm{f}}(\mathrm{G}_K, V) \longrightarrow \mathrm{H}^1_{\mathrm{st}}(\mathrm{G}_K, V) \longrightarrow \mathrm{im}(\beta)/\mathrm{im}(N) \longrightarrow 0.$$

Furthermore, the description of $\mathbf{B}^0_{\mathrm{st}}$ as a pullback allows us to describe

$$\mathrm{im}(\beta)/\mathrm{im}(N) = \big(\mathrm{D}_{\mathrm{st}}(V)/N\mathrm{D}_{\mathrm{st}}(V)\big)^{p\varphi=1}.$$

Since D_{st} commutes with Tate twists and duals, we can also rewrite this as:

$$\begin{aligned}\big(\mathrm{D}_{\mathrm{st}}(V)/N\mathrm{D}_{\mathrm{st}}(V)\big)^{p\varphi=1} &= \big(\mathrm{D}_{\mathrm{st}}(V(-1))/N\mathrm{D}_{\mathrm{st}}(V(-1))\big)^{\varphi=1} \\ &= \big(\mathrm{D}_{\mathrm{st}}(V^\vee(1))\big)^{\varphi=1, N=0} = \big(\mathrm{D}_{\mathrm{cris}}(V^\vee(1))\big)^{\varphi=1}.\end{aligned}$$

This means we can measure on the level of cohomology classes the space of semistable non-crystalline extensions, using the exact sequence

$$0 \longrightarrow \mathrm{H}^1_{\mathrm{f}}(\mathrm{G}_K, V) \longrightarrow \mathrm{H}^1_{\mathrm{st}}(\mathrm{G}_K, V) \longrightarrow \big(\mathrm{D}_{\mathrm{cris}}(V^\vee(1))\big)^{\varphi=1} \longrightarrow 0.$$

This agrees with the previous computation of the dimensions in the case that V is moreover assumed crystalline, so that $\dim_{\mathbb{Q}_p} \big(\mathrm{D}_{\mathrm{cris}}(V^\vee(1))\big)^{\varphi=1} = \dim_{\mathbb{Q}_p} \big(\mathrm{D}_{\mathrm{st}}(V)\big)^{p\varphi=1} = \mathrm{h}^2_{\mathrm{st}}(\mathrm{G}_K, V)$.

Note also that we can use these results to prove that when V is semistable, we have the equality

$$\mathrm{H}^1_{\mathrm{st}}(\mathrm{G}_K, V) = \mathrm{H}^1_{\mathrm{g}}(\mathrm{G}_K, V).$$

In other words: de Rham extensions of semistable representations are necessarily semistable. This was proved in an unpublished paper of Hyodo; see also [Nekovář. Heights, Proposition 1.24].

One application is to consider the case where V is a 1-dimensional semistable (a fortiori, crystalline) $\mathbb{Q}_p$-representation of G_K. In this case, we have:

$$\mathrm{H}^1_{\mathrm{st}}(\mathrm{G}_K, V)/\mathrm{H}^1_{\mathrm{f}}(\mathrm{G}_K, V) \cong \big(\mathrm{D}_{\mathrm{cris}}(V^\vee(1))\big)^{\varphi=1} = \begin{cases} \mathbb{Q}_p & \text{if } V \cong \mathbb{Q}_p(1), \\ 0 & \text{otherwise.} \end{cases}$$

16.1.1 Examples

We just saw that, if we are given a non-split semistable non-crystalline extension of crystalline p-adic Galois representations

$$0 \to \mathbb{Q}_p(\mu) \to W \to \mathbb{Q}_p \to 0,$$

then necessarily $\mu = \chi_p$ is the cyclotomic character. However, this is no longer true if we work with extensions of E-representations, for $E/\mathbb{Q}_p$ some finite extension. Indeed, we will now construct an example of a non-split semistable non-crystalline extension

$$0 \to E(\mu) \to W \to E \to 0$$

for some crystalline character $\mu \neq \chi_p$.

Let $K = K_0$ be an unramified quadratic extension of $\mathbb{Q}_p$, and $E/\mathbb{Q}_p$ a (large enough) coefficient field, which we take to be K_0 itself for convenience. Write also $\mathrm{Gal}\left(K_0/\mathbb{Q}_p\right) = \langle\sigma\rangle$.
Define D, a 2-dimensional filtered (φ, N)-module over K with E-action, as follows. Using the isomorphism $K_0 \otimes_{\mathbb{Q}_p} E \underset{\sim}{\overset{(\tau_1,\tau_2)}{\hookrightarrow}} E \times E$, the underlying (φ, N)-module of D is the free $K_0 \otimes_{\mathbb{Q}_p} E$-module of rank 2 given by

$$D = \overbrace{Ee_{11} \oplus Ee_{12}}^{\tau_1} \oplus \overbrace{Ee_{21} \oplus Ee_{22}}^{\tau_2} = e_1 D \oplus e_2 D,$$

with matrices for φ and N, in the E-basis $(e_{11}, e_{12}, e_{21}, e_{22})$, given as follows:

$$\varphi : \left(\begin{array}{cc|cc} 0 & 0 & pa_1 & 0 \\ 0 & 0 & x_1 & a_1 \\ \hline pa_2 & 0 & 0 & 0 \\ x_2 & a_2 & 0 & 0 \end{array}\right), \qquad N : \left(\begin{array}{cc|cc} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right).$$

Here $a_1, a_2, x_1, x_2 \in K_0$ (with a_1, x_1 considered as elements of E through τ_1, and a_2, x_2 through τ_2), with the requirements:

- $a_1\sigma(a_2) = \sigma(a_1)a_2$,
- $v_p(a_1\sigma(a_2)) = 1$,
- $a_1\sigma(x_2) + p\sigma(a_2)x_1 = 0$.

For instance, one can take $a_1 = p$, $a_2 = 1$, $x_1 = -\sigma(x_2)$.
The filtration on D is given as follows:

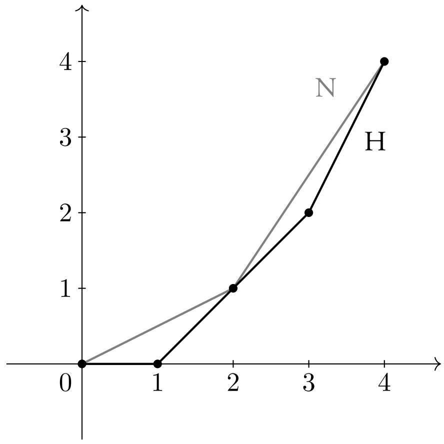
$$\begin{tikzcd} F^3D \arrow[d, equal] \arrow[r, phantom, "\overbrace{\qquad}^{\tau_2}"] & F^2D \arrow[d, equal] \arrow[r, phantom, "\overbrace{\qquad\qquad}^{2\cdot\tau_1}"] & F^1D \arrow[d, equal] \arrow[r, phantom, "\overbrace{\qquad}^{\tau_2}"] & F^0D \arrow[d, equal] \\ 0 \arrow[r, hook] & \langle e_{21} + \alpha e_{22}\rangle \arrow[r, hook] & \langle e_{21} + \alpha e_{22}, e_{11}, e_{22}\rangle \arrow[r, hook] & D \end{tikzcd}$$

for some $\alpha \in K_0$ (embedded through τ_2).
The labelled Hodge–Tate weights of D are thus $\{\overbrace{(1,1)}^{\tau_1}, \overbrace{(0,2)}^{\tau_2}\}$. As for the Newton slopes, we can compute:

$$\varphi^2 : \left(\begin{array}{cc|cc} p^2a_1\sigma(a_2) & 0 & 0 & 0 \\ 0 & a_1\sigma(a_2) & 0 & 0 \\ \hline 0 & 0 & p^2\sigma(a_1)a_2 & 0 \\ 0 & 0 & 0 & \sigma(a_1)a_2 \end{array}\right).$$

Thus the Newton slopes are $\frac{1}{2}$ and $\frac{3}{2}$ (both to be counted with multiplicity $[E : \mathbb{Q}_p] = 2$).

This gives the following Newton and Hodge polygons:



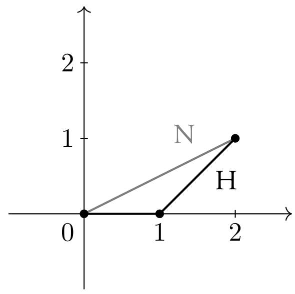
Let D' be a submodule of D. If $D' \not\subseteq \ker(N)$, then $D' = D$ as φ isomorphically maps between $e_1 D$ and $e_2 D$.
For $D' = \ker(N) = Ee_{12} \oplus Ee_{22}$, φ is given by

$$\varphi : \left(\begin{array}{c|c} 0 & a_1 \\ \hline a_2 & 0 \end{array}\right), \qquad \varphi^2 : \left(\begin{array}{c|c} a_1\sigma(a_2) & 0 \\ \hline 0 & \sigma(a_1)a_2 \end{array}\right).$$

The filtration $F^\bullet D' = D' \cap F^\bullet D$ is given by:

$$\begin{tikzcd} F^2 D' \arrow[r, phantom, "\overbrace{\qquad}^{\tau_1}"] \arrow[d, equal] & F^1 D' \arrow[r, phantom, "\overbrace{\qquad}^{\tau_2}"] \arrow[d, equal] & F^0 D' \arrow[d, equal] \\ 0 \arrow[r, hook] & \langle e_{12} \rangle \arrow[r, hook] & D' \end{tikzcd}$$

Therefore the labelled Hodge–Tate weights of $D' = \ker(N)$ are $\{\overbrace{\ 1\ }^{\tau_1}, \overbrace{\ 0\ }^{\tau_2}\}$, and the Newton slope is $\frac{1}{2}$ (with multiplicity 2), giving the following Newton and Hodge polygons:



Let's now show that D' is admissible. Given a submodule of $\ker(N)$, to write down "eigenvectors" of φ, we need some $\beta \in K_0$ with $\mathbf{N}^{K_0}_{\mathbb{Q}_p} \beta = \frac{a_1}{a_2}$. However, $\mathrm{v}_p\left(\frac{a_1}{a_2}\right) = \mathrm{v}_p\left(\frac{a_1\sigma(a_2)}{a_2\sigma(a_2)}\right) = 1 - \mathrm{v}_p(a_2\sigma(a_2))$ is odd, so there is no such β in K_0 and thus $\ker(N)$ has no non-trivial proper submodules,[64] wherefore $D' = \ker(N)$ is admissible and irreducible, so that D is admissible and reducible, an extension of $D'' = D/D'$ by D'.
Twisting by $(D'')^\vee$, we define $\widetilde{D} = D \otimes (D'')^\vee$, $\widetilde{D}' = D' \otimes (D'')^\vee$.

[64] For more general K, there will be non-trivial proper submodules, but they will have $\mathrm{t_N} > \mathrm{t_H}$, so that D' remains admissible and irreducible.

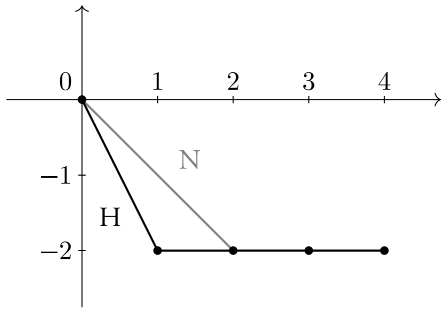
A presentation for $\widetilde{D}$ is as follows:

$$\varphi : \left(\begin{array}{cc|cc} 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{x_1}{pa_1} & \frac{1}{p} \\ \hline 1 & 0 & 0 & 0 \\ \frac{x_2}{pa_2} & \frac{1}{p} & 0 & 0 \end{array}\right), \qquad N : \left(\begin{array}{cc|cc} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right).$$

The filtration on $\widetilde{D}$ is given by:

$$\begin{tikzcd} F^1\widetilde{D} \arrow[r, dash, bend left, "2\cdot\tau_1+\tau_2"] \arrow[d, equal] & F^0\widetilde{D} \arrow[rr, equal] \arrow[dr, equal] & & F^{-1}\widetilde{D} \arrow[r, dash, bend left, "\tau_2"] \arrow[dl, equal] & F^{-2}\widetilde{D} \arrow[d, equal] \\ 0 \arrow[rr, hookrightarrow] & & \langle e_{21} + \alpha e_{22}, e_{11}, e_{12}\rangle \arrow[rr, hookrightarrow] & & \widetilde{D} \end{tikzcd}$$

The labelled Hodge–Tate weights of $\widetilde{D}$ are then $\{\overbrace{(0,0)}^{\tau_1}, \overbrace{(-2,0)}^{\tau_2}\}$, and the Newton and Hodge polygons of $\widetilde{D}$ look like:



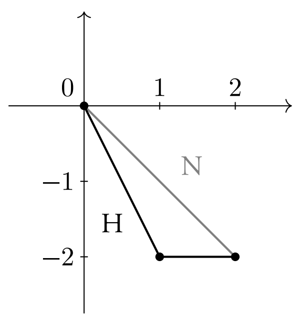
As for $\widetilde{D}'$, we can read off that it is given by

$$\varphi : \left(\begin{array}{c|c} 0 & 1/p \\ \hline 1/p & 0 \end{array}\right), \qquad \varphi^2 : \left(\begin{array}{c|c} 1/p^2 & 0 \\ \hline 0 & 1/p^2 \end{array}\right)$$

with filtration

$$\begin{tikzcd} F^1\widetilde{D} \arrow[r, dash, bend left, "\tau_1"] \arrow[d, equal] & F^0\widetilde{D} \arrow[rr, equal] \arrow[dr, equal] & & F^{-1}\widetilde{D} \arrow[r, dash, bend left, "\tau_2"] \arrow[dl, equal] & F^{-2}\widetilde{D} \arrow[d, equal] \\ 0 \arrow[rr, hookrightarrow] & & \langle e_{12}\rangle \arrow[rr, hookrightarrow] & & \widetilde{D} \end{tikzcd}$$

determining the labelled Hodge–Tate weights as $\{\overbrace{\ 0\ }^{\tau_1}, \overbrace{\ -2\ }^{\tau_2}\}$, and associated Newton and Hodge polygons



Using Proposition 15.2 and the subsequent remark, we can in fact recognise $\widetilde{D}'$ as $\mathrm{D}_{\mathrm{cris}}\left(E(\chi_{\tau_2}^2)\right)$, for χ_{τ_2} the Lubin–Tate character associated with the Lubin–Tate formal group $\Gamma_{K,p}$ and the embedding $\tau_2 \colon K \hookrightarrow E$.

Finally, we can compute $\mathrm{H}^1_{\mathrm{st}}(\mathrm{G}_K, E(\chi_{\tau_2}^2))/\mathrm{H}^1_{\mathrm{f}}(\mathrm{G}_K, E(\chi_{\tau_2}^2)) \cong \left((\widetilde{D}')^\vee(1)\right)^{\varphi=1}$. To do so, it suffices to write out the matrix of φ acting on $(\widetilde{D}')^\vee(1)$; it is the following:

$$\varphi : \left(\begin{array}{c|c} 0 & 1 \\ \hline 1 & 0 \end{array}\right),$$

ergo $\big((\widetilde{D}')^\vee(1)\big)^{\varphi=1} = E$.
To summarise, then, let $W = \mathrm{V}_{\mathrm{st}}(\widetilde{D})$. This is a semistable non-crystalline E-representation of G_K, extension

$$0 \longrightarrow E(\chi_{\tau_2}^2) \longrightarrow W \longrightarrow E \longrightarrow 0$$

of the trivial E-representation by the character $\chi_{\tau_2}^2$. The above calculation

$$\mathrm{H}^1_{\mathrm{st}}(\mathrm{G}_K, E(\chi_{\tau_2}^2))/\mathrm{H}^1_{\mathrm{f}}(\mathrm{G}_K, E(\chi_{\tau_2}^2)) \cong E$$

shows that W is the unique such extension, up to scalars and up to Baer sum with the (unique up to scalars) non-split crystalline extension

$$0 \longrightarrow E(\chi_{\tau_2}^2) \longrightarrow U \longrightarrow E \longrightarrow 0.$$

Bibliography for part I

[André. $F^{\geqslant\lambda}$] Yves André. "Slope filtrations". In: *Confluentes Math.* 1.1 (2009), pp. 1–85.

[Berthelot. $\mathrm{H}_{\mathrm{cris}}$] Pierre Berthelot. *Cohomologie cristalline des schémas de caractéristique* $p > 0$. Lecture Notes in Mathematics, Vol. 407. Springer-Verlag, Berlin-New York, 1974, p. 604.

[Berthelot–Ogus] Pierre Berthelot and Arthur Ogus. *Notes on crystalline cohomology*. Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo, 1978, pp. vi+243.

[Brinon–Conrad] Olivier Brinon and Brian Conrad. *Notes on p-adic Hodge Theory*. 2009. URL: http://math.stanford.edu/~conrad/papers/notes.pdf.

[Cartan–Eilenberg] Henri Cartan and Samuel Eilenberg. *Homological algebra*. Princeton Landmarks in Mathematics. With an appendix by David A. Buchsbaum, Reprint of the 1956 original. Princeton University Press, Princeton, NJ, 1999, pp. xvi+390.

[Colmez. $\int$] Pierre Colmez. "Intégration sur les variétés p-adiques". In: *Astérisque* 248 (1998), pp. viii+155.

[Colmez. $\mathbf{B}_{\mathrm{dR}}^{+}$] Pierre Colmez. "Une construction de $\mathbf{B}_{\mathrm{dR}}^{+}$". In: *Rend. Semin. Mat. Univ. Padova* 128 (2012), 109–130 (2013).

[Colmez–Fontaine] Pierre Colmez and Jean-Marc Fontaine. "Construction des représentations p-adiques semi-stables". In: *Invent. Math.* 140.1 (2000), pp. 1–43.

[Conrad. Local χ] Brian Conrad. "Lifting global representations with local properties". Revision of "Grunwald–Wang for global character groups". Dec. 2011.

[Deligne. $(\boldsymbol{T}, \otimes, \omega)$] Pierre Deligne. "Catégories Tannakiennes". In: *The Grothendieck Festschrift, Vol. II*. Vol. 87. Progress in Mathematics. Birkhäuser Boston, 1990, pp. 111–195.

[Deligne. ∇] Pierre Deligne. *Équations différentielles à points singuliers réguliers*. Lecture Notes in Mathematics, Vol. 163. Springer-Verlag, Berlin-New York, 1970, pp. iii+133.

[Deligne. Hodge II] Pierre Deligne. "Théorie de Hodge. II". In: *Inst. Hautes Études Sci. Publ. Math.* 40 (1971), pp. 5–57.

[Deligne. Hodge III] Pierre Deligne. "Théorie de Hodge. III". In: *Inst. Hautes Études Sci. Publ. Math.* 44 (1974), pp. 5–77.

[Deligne. Weil I] Pierre Deligne. "La conjecture de Weil. I". In: *Inst. Hautes Études Sci. Publ. Math.* 43 (1974), pp. 273–307.

[Deligne. Weil II] Pierre Deligne. "La conjecture de Weil. II". In: *Inst. Hautes Études Sci. Publ. Math.* 52 (1980), pp. 137–252.

[EGA IV_4] Alexander Grothendieck. "Éléments de géométrie algébrique. IV. Étude locale des schémas et des morphismes de schémas (Quatrième Partie)". In: *Inst. Hautes Études Sci. Publ. Math.* 32 (1967). Rédigé avec la collaboration de Jean Dieudonné, p. 361.

[Fontaine. $\mathbf{B}_{\mathrm{dR}}$] Jean-Marc Fontaine. "Le corps des périodes p-adiques". In: *Astérisque* 223 (1994). With an appendix by Pierre Colmez, Périodes p-adiques (Bures-sur-Yvette, 1988), pp. 59–111.

[Fontaine. $\mathbf{B}_{\mathrm{cris}}$] Jean-Marc Fontaine. "Sur certains types de représentations p-adiques du groupe de Galois d'un corps local; construction d'un anneau de Barsotti-Tate". In: *Ann. of Math. (2)* 115.3 (1982), pp. 529–577.

[Fontaine–Perrin-Riou] Jean-Marc Fontaine and Bernadette Perrin-Riou. “Autour des conjectures de Bloch et Kato: cohomologie galoisienne et valeurs de fonctions L”. In: *Motives (Seattle, WA, 1991)*. Vol. 55. Proc. Sympos. Pure Math. Amer. Math. Soc., Providence, RI, 1994, pp. 599–706.

[Grothendieck. Crystals] Alexander Grothendieck. “Crystals and the de Rham cohomology of schemes”. In: *Dix exposés sur la cohomologie des schémas*. Vol. 3. Adv. Stud. Pure Math. Notes by I. Coates and O. Jussila. North-Holland, Amsterdam, 1968, pp. 306–358.

[Hironaka. Resolutions] Heisuke Hironaka. “Resolution of singularities of an algebraic variety over a field of characteristic zero. I, II”. In: *Ann. of Math. (2) 79 (1964), 109–203; ibid. (2)* 79 (1964), pp. 205–326.

[Hyodo–Kato] Osamu Hyodo and Kazuya Kato. “Semi-stable reduction and crystalline cohomology with logarithmic poles”. In: *Astérisque* 223 (1994). Périodes p-adiques (Bures-sur-Yvette, 1988), pp. 221–268.

[Illusie. $N \circlearrowright V/\mathbf{Q}_\ell$] Luc Illusie. “Autour du théorème de monodromie locale”. In: *Astérisque* 223 (1994). Périodes p-adiques (Bures-sur-Yvette, 1988), pp. 9–57.

[Illusie. $\mathrm{W}_\star\Omega^\bullet$] Luc Illusie. “Complexe de de Rham-Witt et cohomologie cristalline”. In: *Ann. Sci. École Norm. Sup. (4)* 12.4 (1979), pp. 501–661.

[Illusie. $\mathrm{H}_{\mathrm{cris}}$] Luc Illusie. “Crystalline cohomology”. In: *Motives (Seattle, WA, 1991)*. Vol. 55. Proc. Sympos. Pure Math. Amer. Math. Soc., Providence, RI, 1994, pp. 43–70.

[Ito. W-M in char. p] Tetsushi Ito. “Weight-monodromy conjecture over equal characteristic local fields”. In: *Amer. J. Math.* 127.3 (2005), pp. 647–658.

[de Jong. Alterations] Aise Johan de Jong. “Smoothness, semi-stability and alterations”. In: *Inst. Hautes Études Sci. Publ. Math.* 83 (1996), pp. 51–93.

[Joyal. δ-rings] André Joyal. “δ-anneaux et vecteurs de Witt”. In: *C. R. Math. Rep. Acad. Sci. Canada* 7.3 (1985), pp. 177–182.

[K. Kato. Log structures] Kazuya Kato. “Logarithmic structures of Fontaine-Illusie”. In: *Algebraic analysis, geometry, and number theory (Baltimore, MD, 1988)*. Johns Hopkins Univ. Press, Baltimore, MD, 1989, pp. 191–224.

[Katz–Oda] Nicholas M. Katz and Tadao Oda. “On the differentiation of de Rham cohomology classes with respect to parameters”. In: *J. Math. Kyoto Univ.* 8 (1968), pp. 199–213.

[Kedlaya. Weil II] Kiran S. Kedlaya. “Fourier transforms and p-adic ‘Weil II’”. In: *Compos. Math.* 142.6 (2006), pp. 1426–1450.

[Lubin–Tate] Jonathan Lubin and John Tate. “Formal complex multiplication in local fields”. In: *Ann. of Math. (2)* 81 (1965), pp. 380–387.

[Milgram–Rosenbloom] Arthur Milgram and Paul Rosenbloom. “Harmonic forms and heat conduction. I. Closed Riemannian manifolds”. In: *Proc. Nat. Acad. Sci. U. S. A.* 37 (1951), pp. 180–184.

[Nagata. $X \hookrightarrow \bar{X}$] Masayoshi Nagata. “A generalization of the imbedding problem of an abstract variety in a complete variety”. In: *J. Math. Kyoto Univ.* 3 (1963), pp. 89–102.

[Nakkajima. ${}_W\mathrm{E}_1^{p,q}$] Yukiyoshi Nakkajima. “p-adic weight spectral sequences of log varieties”. In: *J. Math. Sci. Univ. Tokyo* 12.4 (2005), pp. 513–661.

[Nekovář. Heights] Jan Nekovář. “On p-adic height pairings”. In: *Séminaire de Théorie des Nombres, Paris, 1990–91*. Vol. 108. Progr. Math. Birkhäuser Boston, Boston, MA, 1993, pp. 127–202.

[Ogus. Log notes] Arthur Ogus. *Lectures on Logarithmic Algebraic Geometry*. 2006. URL: https://math.berkeley.edu/~ogus/preprints/log_book/logbook.pdf.

[Rapoport–Zink] Michael Rapoport and Thomas Zink. “Über die lokale Zetafunktion von Shimuravarietäten. Monodromiefiltration und verschwindende Zyklen in ungleicher Charakteristik”. In: *Invent. Math.* 68.1 (1982), pp. 21–101.

[M. Saito. $\mathrm{MH}^{\mathrm{pol}}$] Morihiko Saito. “Modules de Hodge polarisables”. In: *Publ. Res. Inst. Math. Sci.* 24.6 (1988), 849–995 (1989).

[T. Saito. $_W\mathrm{E}_1^{p,q}$] Takeshi Saito. "Weight spectral sequences and independence of ℓ". In: *J. Inst. Math. Jussieu* 2.4 (2003), pp. 583–634.

[Schmid. Variation] Wilfried Schmid. "Variation of Hodge structure: the singularities of the period mapping". In: *Invent. Math.* 22 (1973), pp. 211–319.

[SGA 4_{II}] *Théorie des topos et cohomologie étale des schémas. II*. Lecture Notes in Mathematics, Vol. 270. Séminaire de Géométrie Algébrique du Bois-Marie 1963–1964 (SGA 4). Dirigé par Michael Artin, Alexander Grothendieck et Jean-Louis Verdier. Avec la collaboration de Pierre Deligne et Bernard Saint-Donat. Springer-Verlag, Berlin-New York, 1972, pp. iv+418.

[SGA 7_{II}] *Groupes de monodromie en géométrie algébrique. II*. Lecture Notes in Mathematics, Vol. 340. Séminaire de Géométrie Algébrique du Bois-Marie 1967–1969 (SGA 7). Dirigé par Pierre Deligne et Nick Katz. Springer-Verlag, Berlin-New York, 1973, pp. x+438.

[Steenbrink. lim $F^\bullet$] Joseph Steenbrink. "Limits of Hodge structures". In: *Invent. Math.* 31.3 (1975/76), pp. 229–257.

[Totaro. $D \otimes D'$] Burt Totaro. "Tensor products in p-adic Hodge theory". In: *Duke Math. J.* 83.1 (1996), pp. 79–104.

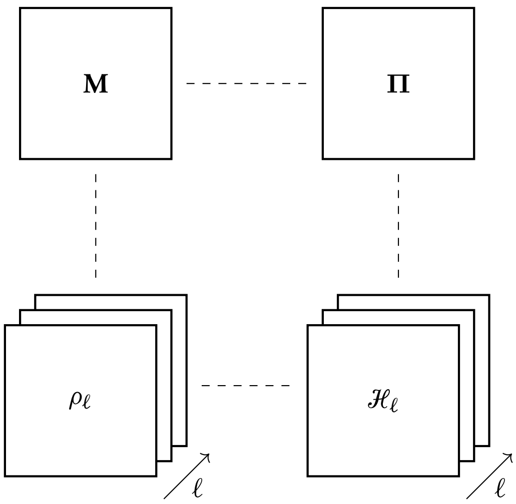
[Tsuji. C_{st}] Takeshi Tsuji. "p-adic étale cohomology and crystalline cohomology in the semi-stable reduction case". In: *Invent. Math.* 137.2 (1999), pp. 233–411.

[Tsuji. C_{st} (survey)] Takeshi Tsuji. "Semi-stable conjecture of Fontaine–Jannsen: a survey". In: *Astérisque* 279 (2002). Cohomologies p-adiques et applications arithmétiques, II, pp. 323–370.

II

Automorphic representations

To inscribe the results of part I in a greater context, we invoke the Langlands philosophy.
On the geometric side, we were dealing with local Galois representations $\rho\colon \mathrm{G}_K \to E$, for finite extensions $K/\mathbb{Q}_p$ and $E/\mathbb{Q}_\ell$ (most of the time with $\ell = p$, as this is the case of interest in p-adic Hodge theory). One particular source of such representations is given by taking local components of global Galois representations $\rho\colon \mathrm{G}_F \to E$, for some number field F (and E as above). Of particular interest are those global Galois representations of geometric origin: those arising in some way from the cohomology of schemes. These schemes (or, more generally, motives) give rise to compatible systems of Galois representations $\{\rho_\ell\}_\ell$, unramified at almost all primes and de Rham at places above ℓ.
On the other side are automorphic objects, such as modular forms, or more generally automorphic representations, which are certain representations of $\boldsymbol{G}(\mathbb{A}_F)$, for $\boldsymbol{G}$ some reductive group over F (and $\mathbb{A}_F$ denoting the adèles of F). Not all automorphic representations π are of interest: this time it is important to impose algebraicity conditions on π_∞.
We thus expect a picture resembling the following:



In brief:

- **M**: motives over F.
- **Π**: algebraic automorphic representations of reductive groups over F.
- $\{\rho_\ell\}_\ell$: collections of ℓ-adic Galois representations $\rho_\ell\colon \mathrm{G}_F \to E$, $E/\mathbb{Q}_\ell$.
- $\{\mathcal{H}_\ell\}_\ell$: collections of representations of Hecke algebras.

1 Langlands' philosophy by example

We quickly sketch some elements pertaining to the Langlands philosophy in the case of modular forms.
Let us start with the simplest example of a cuspidal modular Hecke eigenform of nontrivial level, the Eichler modular form $f \in \mathrm{S}_2(\Gamma_0(11))$, $f(z) = \eta(z)^2\eta(11z)^2$, with η denoting the Dedekind eta function. The Eichler modular form f has q-expansion

$$\begin{aligned} f(z) &= q\prod_{n=1}^{+\infty}(1-q^n)^2(1-q^{11n})^2 \\ &= q - 2q^2 - q^3 + 2q^4 + q^5 + 2q^6 - 2q^7 - 2q^9 - 2q^{10} + q^{11} + \cdots, \end{aligned}$$

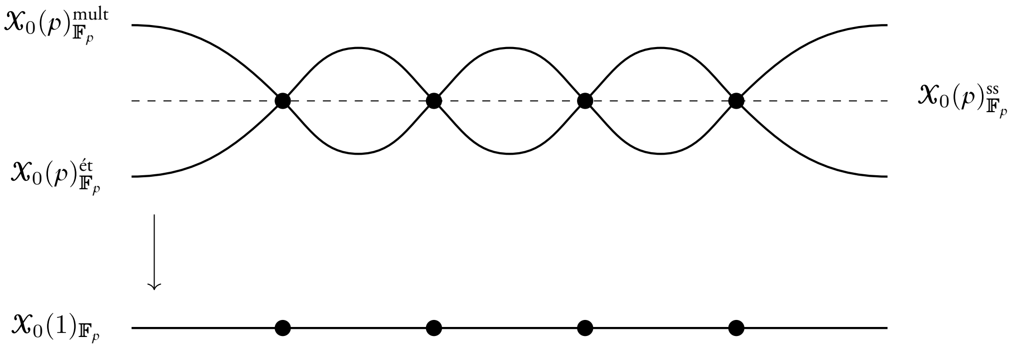
where $q = \exp(2\pi i z)$.

The natural geometric habitat for this modular form f is the modular curve $\mathrm{X}_0(11)$ (a smooth projective curve over $\mathbb{Q}$) and we can naturally view f as an element of $\mathrm{H}^0\left(\mathrm{X}_0(11), \Omega^1_{\mathrm{X}_0(11)/\mathbb{Q}}\right)$; in fact f spans this $\mathbb{Q}$-vector space, because $\mathrm{X}_0(11)$ has genus 1.
Now, we can invoke Hodge theory: $\mathrm{H}^0\left(\mathrm{X}_0(11)_{\mathbb{C}}, \Omega^1_{\mathrm{X}_0(11)_{\mathbb{C}}/\mathbb{C}}\right)$ is naturally identified with $F^1\mathrm{H}^1_{\mathrm{dR}}(\mathrm{X}_0(11)_{\mathbb{C}}/\mathbb{C})$, so that by oppositeness f and its complex conjugate $\bar{f}$ form a basis of the complex vector space $\mathrm{H}^1_{\mathrm{dR}}(\mathrm{X}_0(11)_{\mathbb{C}}/\mathbb{C})$.
This suggests a location for Galois representations: the ℓ-adic cohomology groups $\mathrm{H}^1_{\text{ét}}\left(\mathrm{X}_0(11)_{\overline{\mathbb{Q}}}, \mathbb{Q}_\ell\right)$. For any prime number ℓ, write $\rho_{f,\ell}\colon \mathrm{G}_{\mathbb{Q}} \to \mathrm{Aut}\left(\mathrm{H}^1_{\text{ét}}\left(\mathrm{X}_0(11)_{\overline{\mathbb{Q}}}, \mathbb{Q}_\ell\right)\right)$ for the action by transport of structure of $\mathrm{G}_{\mathbb{Q}}$ on the ℓ-adic cohomology of $\mathrm{X}_0(11)$; this is a 2-dimensional $\mathbb{Q}_\ell$-representation by the above.

We can in fact be even more explicit in our description of these representations. The curve $\mathrm{X}_0(11)$ corresponds to the Weierstrass equation

$$\mathfrak{W}: \; y^2 + y = x^3 - x^2 - 10x - 20,$$

with j-invariant $\mathrm{j} = -2^{12}11^{-5}31^3$ and discriminant $\Delta = -11^5$ (which is minimal). This means that the Weierstrass model $\mathfrak{W}$ is a flat proper scheme over $\mathbb{Z}$, smooth over $\mathbb{Z}[1/11]$. In particular, the representation $\rho_{f,\ell}$ is unramified at p for any $p \neq 11$ with $p \neq \ell$ (respectively, crystalline at p if $p = \ell \neq 11$). The behaviour that is left to understand is thus for $p = 11$.
Another way to tackle the integral structure of $\mathrm{X}_0(11)$ is by using a moduli interpretation. The open modular curve $\mathrm{Y}_0(N)$ classifies elliptic curves with a given cyclic subgroup of order N. The compactification $\mathrm{X}(1)$ of $\mathrm{Y}(1)$ corresponds to allowing semistable degenerations, in the form of a nodal cubic curve. We have a similar description for $\mathrm{X}_0(N)$ [Deligne–Rapoport, Théorème V.1.6] : $\mathrm{X}_0(N)$ classifies generalised elliptic curves E equipped with a subgroup scheme C, locally free of rank N, which meets every irreducible component of every geometric fibre of E.
It is then possible to use these moduli problems to study the integral models. For instance, we have the following explicit description of the integral model $\mathcal{X}_0(p)$ over $\mathbb{Z}$ [Deligne–Rapoport, Théorème V.1.16] : $\mathcal{X}_0(p)_{\mathbb{F}_p}$ corresponds to two copies of $\mathcal{X}(1)_{\mathbb{F}_p}$ glued together transversally. The idea is to map an elliptic curve $E/\mathbb{F}_p$ (corresponding to a point in $\mathcal{X}(1)(\mathbb{F}_p)$) to the points $(E, \ker(F)\colon E \to E^{(p)})$ $(E^{(p)}, \ker(V)\colon E^{(p)} \to E)$; these correspond to distinct points in $\mathcal{X}_0(p)(\mathbb{F}_p)$ unless E is supersingular. Hence we can obtain $\mathcal{X}_0(p)_{\mathbb{F}_p}$ by gluing two copies of $\mathcal{X}(1)_{\mathbb{F}_p}$ along supersingular points, identifying the point $[E]$ of the first copy to the point $\left[E^{(p)}\right]$ in the second copy. We can visualise this as follows:



In particular, this means that $\mathcal{X}_0(p)_{\mathbb{Z}_p}$ is a semistable scheme. The local Galois representation $\rho_{f,\ell,11}\colon \mathrm{G}_{\mathbb{Q}_{11}} \to \mathrm{H}^1_{\text{ét}}\left(\mathcal{X}_0(11)_{\overline{\mathbb{Q}_{11}}}, \mathbb{Q}_\ell\right)$ is thus semistable (which means that, for $\ell \neq 11$, the inertia subgroup $\mathrm{I}_{11} \subset \mathrm{G}_{\mathbb{Q}_{11}}$ acts unipotently; and for $\ell = 11$, that $\rho_{f,\ell,11}$ is $\mathbf{B}_{\text{st}}$-admissible). We also know that $\rho_{f,\ell,11}$ is not unramified, and $\rho_{f,11,11}$ is not crystalline.

Another example will serve to illustrate some of the complications that can appear. Consider now $\mathrm{S}_8(\Gamma_0(9))$. As a module for the Hecke algebra, the subspace of newforms splits as the direct sum of two subspaces. One is 1-dimensional over $\mathbb{Q}$, spanned by the Hecke eigenform

$$f_1 : q - 6q^2 - 92q^4 - 390q^5 - 64q^7 + 1320q^8 + 2340q^{10} + \cdots,$$

while the other is two dimensional over $\mathbb{Q}$, and is spanned by the Hecke eigenform

$$f_2 : q + 6\alpha q^2 + 232q^4 - 96\alpha q^5 + 260q^7 + 624\alpha q^8 - 5760q^{10} + \cdots,$$

where $\alpha = \sqrt{10}$. On top of these newforms, we get two contributions from $\mathrm{S}_8(\Gamma_0(3))$, which is spanned by the form

$$g : q + 6q^2 – 27q^3 – 92q^4 + 390q^5 – 162q^6 – 64q^7 – 1320q^8 + 729q^9 + 2340q^{10} + \cdots.$$

Associated to $g \in \mathrm{S}_8(\Gamma_0(3))$ are the two oldforms $g_1 = i_1(g)$, with the same q-expansion as g, and $g_2 = i_3(g)$, with q-expansion

$$g_2 : q^3 + 6q^6 – 27q^9 – 92q^{12} + \cdots.$$

This time, in weight 8, we will find the above cusp forms in the space of global sections $\mathrm{H}^0\left(\mathrm{X}_0(9), \Omega^1_{\mathrm{X}_0(9)/\mathbb{Q}} \otimes_{\mathcal{O}_{\mathrm{X}_0(9)}} \omega^{\otimes 6}_{\mathrm{X}_0(9)}\right)$. For conciseness, we shall write

$$\omega(k) := \omega^{\otimes k}_{\mathrm{X}_0(N)}, \qquad \omega_0(k) := \Omega^1_{\mathrm{X}_0(N)/\mathbb{Q}} \otimes_{\mathcal{O}_{\mathrm{X}_0(N)}} \omega(k-2).$$

On the Hodge theoretic side, the local system corresponding to $\omega_0(k)$ is $\mathcal{L}(k) := \bigvee^{k-2} \mathrm{Sd}$, where Sd denotes the local system on $\mathrm{X}_0(N)$ given by the standard representation of GL_2 [Deligne. $\rho_{f,\ell}$, 2.5, 2.9]. Introducing the parabolic cohomology groups $\mathrm{H}^\bullet_{\text{dR},!}(-)$, which are the image of the compactly supported cohomology groups $\mathrm{H}^\bullet_{\text{dR},c}(-)$ in $\mathrm{H}^\bullet_{\text{dR}}(-)$,[65] Hodge theory then treats us to the **Shimura isomorphism** [Deligne. $\rho_{f,\ell}$, Théorème 2.10] :

$$\mathrm{H}^0\left(\mathrm{X}_0(N)_{\mathbb{C}}, \omega_0(k)\right) \oplus \overline{\mathrm{H}^0\left(\mathrm{X}_0(N)_{\mathbb{C}}, \omega_0(k)\right)} \xrightarrow{\sim} \mathrm{H}^1_{\text{dR},!}\left(\mathrm{Y}_0(N)/\mathbb{C}, \mathcal{L}(k)\right).$$

In particular, $\mathrm{H}^1_{\text{dR},!}\left(\mathrm{Y}_0(9)/\mathbb{C}, \mathcal{L}(8)\right)$ splits up into 4 pieces: 2-dimensional subspaces attached to each of f_1, g_1, and g_2, and a 4-dimensional subspace attached to f_2.

To get our hands on the ℓ-adic representations, we can also use the standard representation of GL_2 to define ℓ-adic local systems $\mathcal{L}_\ell(k)$. We then want to consider

$$\mathrm{H}^1_{\text{ét},!}\left(\mathrm{Y}_0(9)_{\overline{\mathbb{Q}}}, \mathcal{L}_\ell(8)\right),$$

[65] Alternatively, the groups $\mathrm{H}^\bullet_{\text{dR},!}(Y/\mathbb{C}, \mathcal{F})$ are naturally isomorphic to the L^2-cohomology groups $\mathrm{H}^\bullet_{(2)}(Y(\mathbb{C}), \mathcal{F})$.

which comes equipped with an action of Hecke operators for $\mathrm{GL}_{2,\mathbb{Q}}$, such as T_p. We can then pick out submodules on which these T_p-operators act isotypically. This is less interesting for g_1 and g_2, as they are oldforms and thus can be found in the cohomology of $\mathrm{X}_0(3)$. We concentrate on the two Galois representations

$$\rho_{f_1,\ell}\colon \mathrm{G}_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{Q}_\ell), \quad \text{and} \quad \rho_{f_2,\ell}\colon \mathrm{G}_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{Q}_\ell(\alpha)).$$

However, f_1 is not particularly interesting either, as it is a twist of g_1. Let's now take a better look at f_2, which is both new at 9 and not a twist of a form of lower level. As $\mathrm{X}_0(9)$ is smooth over $\mathbb{Z}[1/3]$, the representations $\rho_{f_2,\ell}$ are determined by the characteristic polynomials of Frob_p for $p \nmid 3\ell$, which are $X^2 - \mathrm{a}_p(f_2)X + p^7$.
We are left with the problem of understanding the local Galois representations $\rho_{f_2,\ell,3}\colon \mathrm{G}_{\mathbb{Q}_3} \to \mathrm{GL}_2(\mathbb{Q}_\ell(\alpha))$. Shifting focus to the automorphic side (refer to section 2 for relevant definitions), this prompts us to consider the automorphic representation π_{f_2} attached to f_2, and in particular its local component at 3, $\pi_{f_2,3}$. This is a smooth admissible irreducible representation of $\mathrm{GL}_2(\mathbb{Q}_3)$ of conductor 3^2 and central character $|-|^{-6}$, which is not a twist of a representation of lower conductor; it is thus supercuspidal [Loeffler–Weinstein, Proposition 2.8]. Writing $U := \mathrm{GL}_2(\mathbb{Z}_3)$, a maximal compact open subgroup of $\mathrm{GL}_2(\mathbb{Q}_3)$, and $U_n := \begin{pmatrix} 1+3^n\mathbb{Z}_3 & 3^n\mathbb{Z}_3 \\ 3^n\mathbb{Z}_3 & 1+3^n\mathbb{Z}_3 \end{pmatrix} \subset \mathrm{GL}_2(\mathbb{Z}_3)$, it then follows that this representation must have a vector fixed under U_1, for otherwise it would have conductor at least 3^3 [Breuil–Mézard, A3.2].[66] This means that $\pi_{f_2,3}$ can be obtained from a cuspidal representation of $U/U_1 \cong \mathrm{GL}_2(\mathbb{F}_3)$ by inflation and induction. More specifically, given an irreducible cuspidal representation $\xi\colon \mathrm{GL}_2(\mathbb{F}_3) \to \mathrm{GL}_d(\mathbb{C})$, we can perform the following steps

- start with $\bar{\xi}\colon U/U_1 \cong \mathrm{GL}_2(\mathbb{F}_3) \to \mathrm{GL}_d(\mathbb{C})$,
- inflate $\bar{\xi}$ to $\xi\colon U \cong \mathrm{GL}_2(\mathbb{Z}_3) \to \mathrm{GL}_d(\mathbb{C})$,
- extend ξ from U to ZU by setting $\tilde{\xi}(zu) = |z|^s\xi(u)$ for some $s \in \mathbb{C}$,[67] where $Z \cong \mathbb{Q}_3^\times$ is the centre of $\mathrm{GL}_2(\mathbb{Q}_3)$,
- induce $\tilde{\xi}$, obtaining the representation $\mathrm{ind}_{ZU}^{\mathrm{GL}_2(\mathbb{Q}_3)}\tilde{\xi}$ of $\mathrm{GL}_2(\mathbb{Q}_3)$.[68]

The representation $\mathrm{ind}_{ZU}^{\mathrm{GL}_2(\mathbb{Q}_3)}\tilde{\xi}$ is then an irreducible supercuspidal representation of $\mathrm{GL}_2(\mathbb{Q}_3)$, and all irreducible supercuspidal representations of $\mathrm{GL}_2(\mathbb{Q}_3)$ which contain an U_1-invariant vector can be obtained thusly.
The question remains: which cuspidal representation ξ of $\mathrm{GL}_2(\mathbb{F}_3)$ recovers $\pi_{f_2,3}$?
To start off, Deligne–Lusztig theory describes the cuspidal representations of $\mathrm{GL}_2(\mathbb{F}_3)$ as "non-split parabolic inductions" (through the process of **Deligne–Lusztig induction**). In fact, $\mathrm{GL}_2(\mathbb{F}_3)$ has $\frac{3(3-1)}{2} = 3$ cuspidal representations, all of dimension $3-1=2$. These correspond by Deligne–Lusztig theory to characters of the unique (up to conjugation) non-split torus T of $\mathrm{GL}_2(\mathbb{F}_3)$. More explicitly, describing $T \cong (\mathbb{F}_{3^2})^\times$ by choosing a generator $\eta = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ of T, the aforementioned cuspidal representations correspond to the three irreducible characters of T which don't factor through the norm map $\mathbb{F}_{3^2} \to \mathbb{F}_3$, up to Galois action: we must simply choose an 8-th root of unity in $\mathbb{C}$ to which we send η, of order at least 4 (because of the norm condition). Up to Frobenius, the three choices are i, $\frac{\sqrt{2}(1+\mathrm{i})}{2}$ and $\frac{\sqrt{2}(1-\mathrm{i})}{2}$.
To figure out which root of unity is relevant to $\pi_{f_2,3}$, we note that the only cuspidal representation of $\mathrm{GL}_2(\mathbb{F}_3)$ of the three above with trivial central character corresponds to mapping η to i. As f_2 has trivial Nebentypus, this is the cuspidal representation ξ we are looking for, such that $\mathrm{ind}_{ZU}^{\mathrm{GL}_2(\mathbb{Q}_3)}\tilde{\xi} \cong \pi_{f_2,3}$.
The Langlands philosophy then leads us to expect that the corresponding local Galois representations $\rho_{f_2,\ell}\colon \mathrm{G}_{\mathbb{Q}_3} \to \mathrm{GL}_2(\mathbb{Q}_\ell(\alpha))$ are potentially unramified (respectively, potentially crystalline) but not unramified

[66] In which case we would have to look at invariants under U_n for $n > 1$, and representations of $\mathrm{GL}_2(\mathbb{Z}/p^n\mathbb{Z})$.
[67] In the case of the local component of a modular form of weight k, we must take $s = 2-k$ to obtain the correct (Hecke compatible) central character.
[68] This is the induction functor, also called compact induction, which is *left* adjoint to the restriction functor. This is distinct from co-induction, which is *right* adjoint to restriction.

(respectively, not crystalline) for $\ell \neq 3$ (respectively, $\ell = 3$), with irreducible associated Weil–Deligne representations (with $N = 0$).

These examples illustrate the general situation. We can pick out the representation associated with a cuspidal eigenform f of weight k in the ℓ-adic étale cohomology of the locally constant sheaf $\mathscr{L}(k)$ on the relevant modular curve. The geometry of this modular curve then dictates the representation-theoretic properties of the local Galois representations $\rho_{f,\ell,p}\colon \mathrm{G}_{\mathbb{Q}_p} \to \mathrm{GL}_2(E_{f,\ell})$; more precisely, we can actually pick out geometrically the relevant object, to get a closer correspondence [Scholl. Motives, Theorem 1.2.4]. This matches up properties of the local component $\pi_{f,p}$ of π_f at p with properties of $\rho_{f,\ell,p}$. In this situation, with the two cases $\ell \neq p$ and $\ell = p$, the Langlands philosophy for instance predicts:

- $\rho_{f,\ell,p}$ is unramified (respectively, crystalline) if and only if $\pi_{f,p}$ is unramified (meaning that $\pi_{f,p}^{\mathrm{GL}_2(\mathbb{Z}_p)} \neq 0$),
- $\rho_{f,\ell,p}$ is semistable if and only if $\pi_{f,p}$ is Iwahori-spherical; that is, $\pi_{f,p}^{I} \neq 0$, where I is the Iwahori subgroup of $\mathrm{GL}_2(\mathbb{Q}_p)$,
- the Weil–Deligne representation attached to $\rho_{f,\ell,p}$ is irreducible if and only if $\pi_{f,p}$ is supercuspidal.

These compatibilities are known, by work of Deligne for $\ell \neq p$ [Deligne. Letter to P-S, §C. Theorem, §D. (A)] (see also [Carayol. $\rho_{f,\ell}$, 11.1], which generalises to the setting of Hilbert modular forms of regular weight), and by work of Saito for $\ell = p$ [T. Saito. HMF, Theorem 1] (also valid for the case of Hilbert modular forms of regular weight).

At this point, we invite the reader to reconsult the second part of the introduction, which recalls the reasons for studying the behaviour of Hecke operators on spaces of Hilbert modular forms, in relation to the questions studied in part I. In particular, we will take interest in studying representations of p-adic algebraic groups which are Iwahori spherical, as these are conjecturally tied, via the Langlands program, to semistable p-adic Galois representations.

2 Representation theory of p-adic groups

The previous examples illustrate the relationship between local Galois representations associated with modular forms and representations of reductive algebraic groups over p-adic fields. Let us quickly review some essential notions pertaining to the latter.

2.1 Smooth admissible representations of p-adic groups

We will be interested in studying the local components $\pi_{f,p}$, with a view towards understanding the Iwahori-spherical representations (as defined in the upcoming section 2.2).

Let $K/\mathbb{Q}_p$ be a p-adic field, and $\boldsymbol{G}/K$ a connected reductive group. We want to study the representation theory of $G = \boldsymbol{G}(K)$, concentrating on representations of G that appear as local factors of automorphic representations.
There are two fundamental notions. Let E be a field of characteristic 0, V a vector space over E, and π a representation of G on V.

- π is **smooth** if the stabiliser of any vector under π is open. Equivalently, π is smooth if the action $(g, v) \mapsto \pi(g)v$ is continuous when V is equipped with the discrete topology.
- π is **admissible** if, for every compact open subgroup U of G, the invariant subspace V^U is finite dimensional.

These are the kind of representations we are interested in. To study them, we can use compact open subgroups to our advantage:

Proposition 2.1 — Let U be a compact open subgroup of G.

- The restriction to U of any smooth representation of G is completely reducible.
- The "U-invariants" functor $(-)^U$ is exact. ◇□

2.2 The Hecke algebra of a p-adic group

We want to know what kind of object V^U is. We say (π, V) is U**-spherical** if every G-stable subspace, of both π and $\tilde{\pi}$, contains a nonzero vector fixed by U. To study U-spherical representations, we define the Hecke algebra $\mathcal{H}(G /\!\!/ U, E)$. This is the E-vector space of compactly supported smooth functions $f : G \to E$ (with E given the discrete topology) that are bi-invariant under U. The algebra structure is given by the convolution product.

Proposition 2.2 — Let U be a compact subgroup of G. The "U-invariants" functor $(-)^U$ induces an equivalence of categories

$$\left\{ \begin{array}{c} U\text{-spherical smooth admissible} \\ E\text{-representations of } G \end{array} \right\} \xrightarrow{(-)^U} \left\{ \text{Smooth } E\text{-representations of } \mathcal{H}(G /\!\!/ U, E) \right\}. \quad ◇□$$

We can also allow U to vary, and define the Hecke algebra $\mathcal{H}(G, E)$ consisting of all compactly supported smooth functions $f : G \to E$. This is again an algebra under convolution, but lacks a unit in general. However, for any compact open subgroup U, we can define the idempotent $e_U = \mu(U)^{-1}\mathbb{1}_U$; we have that $e_U \mathcal{H}(G, E) e_U = \mathcal{H}(G /\!\!/ U, E)$. This allows us to extend the previous proposition, as every vector $v \in V$ is fixed by some compact open subgroup U.

Proposition 2.3 — The "U-invariants" functors, for U ranging over compact subgroups of G, patch together to give an equivalence of categories

$$\left\{ \begin{array}{c} \text{Smooth admissible} \\ E\text{-representations of } G \end{array} \right\} \longrightarrow \left\{ \text{Smooth } E\text{-representations of } \mathcal{H}(G, E) \right\}. \quad ◇□$$

2.3 The Harish-Chandra philosophy

The Harish-Chandra philosophy is a p-adic incarnation of ideas in the representation theory of symmetric groups and finite groups of Lie type, according to which one can construct all representations via processes of induction starting from certain distinguished representations.
The goal is to distill the complexity of the representation theory of reductive algebraic groups over a p-adic field K down to essential ingredients, the **supercuspidal** representations, from which one can obtain the others through parabolic induction.
We shall illustrate this with the **Bernstein–Zelevinsky classification**, which carries out this project in full detail for the case of $\mathrm{GL}_n(K)$, extending Green's classification of complex representations of $\mathrm{GL}_n(\mathbb{F}_q)$.

Recall that representations of the symmetric group can be described by using the induction functors from standard Young subgroups $\mathrm{S}_n \times \mathrm{S}_m$ to S_{n+m}. With general linear groups, the natural analogue of these Young subgroups are the standard Levi subgroups $\mathrm{GL}_n(\mathbb{F}_q) \times \mathrm{GL}_m(\mathbb{F}_q) \subset \mathrm{GL}_{n+m}(\mathbb{F}_q)$ of block diagonal matrices. Instead of performing a direct induction from $\mathrm{GL}_n(\mathbb{F}_q) \times \mathrm{GL}_m(\mathbb{F}_q)$ to $\mathrm{GL}_{n+m}(\mathbb{F}_q)$, it is more natural to start by inflating the representation to the standard parabolic subgroup $P_{n,m}$ before then taking an induction from $P_{n,m}$ to $\mathrm{GL}_{n+m}(\mathbb{F}_q)$.

In the study of representations of symmetric groups, one can immediately notice that all representations can be obtained in this matter, starting with the trivial representation of S_1.
If we do the same for $G = \mathrm{GL}_n(\mathbb{F}_q)$, we obtain the **unipotent** representations of $\mathrm{GL}_n(\mathbb{F}_q)$. Of particular interest is the **Steinberg representation** St, which is the unique irreducible quotient of the parabolic induction $\mathrm{Ind}_B^G \mathbf{1}_B$,[69] where B is the standard Borel subgroup of G, consisting of upper triangular matrices. This is the analogue of the sign representation of S_n.
However, as the nomenclature indicates, the unipotent representations do not exhaust the set of isomorphism classes of all complex representations of $\mathrm{GL}_n(\mathbb{F}_q)$. Instead, there will be several fundamental representations, called **cuspidal representations**, from which the other representations can be obtained by parabolic induction; the trivial representation being only one example of a cuspidal representation. One can in fact use this as a definition: a cuspidal representation is one which is not a subquotient of any parabolically induced representation (from a *proper* parabolic subgroup). This means that, once we understand cuspidal representations, we have a method for constructing all the representations of $\mathrm{GL}_n(\mathbb{F}_q)$.
It is also possible to directly construct these cuspidal representations. The fundamental insight of Lusztig is that, despite cuspidal representations not being parabolically induced, there *is* a way of realising them as twisted parabolic inductions, from parabolic subgroups which are only defined over extensions of $\mathbb{F}_q$. To understand this, it is useful to reformulate parabolic induction by considering parabolically induced representations as those appearing in the cohomology of the flag variety associated with the apposite parabolic subgroup. One can then construct twisted flag varieties, called Deligne–Lusztig varieties, such that cuspidal representations appear in their étale cohomology. This method in fact works for arbitrary groups of Lie type, not just GL_n.

In considering the situation for $G = \mathrm{GL}_n(K)$ for $K/\mathbb{Q}_p$, much of the same phenomena subsist. We can still perform parabolic induction, and have a notion of **supercuspidal representations**, which are those that do not appear as a subquotient of any parabolically induced representation (from a *proper* parabolic subgroup). The Bernstein–Zelevinsky classification takes these representations as given, and aims to understand the structure of general smooth admissible representations of $\mathrm{GL}_n(K)$. The essential ingredient is again the Steinberg representation, which can be defined as above as the unique irreducible quotient of $\mathrm{Ind}_B^G \mathbf{1}_B$.
The Steinberg representation is a representative example of the behaviour of parabolic induction. From the point of view of the local Langlands correspondence, subquotients of $\mathrm{Ind}_B^G \mathbf{1}_B$ correspond to L-parameters $\psi\colon \mathrm{W}_K \times \mathrm{SL}_2(\mathbb{C}) \to \mathrm{GL}_n(\mathbb{C})$ which are trivial on W_K, and differ only on $\mathrm{SL}_2(\mathbb{C})$. For instance, the trivial

[69] This is now (non-normalised) co-induction.

representation is the unique irreducible subrepresentation of $\mathrm{Ind}_B^G \mathbf{1}_B$, and its L-parameter is obviously trivial on $\mathrm{SL}_2(\mathbb{C})$; at the other extreme, the unique irreducible quotient of $\mathrm{Ind}_B^G \mathbf{1}_B$ is the Steinberg representation St, with L-parameter whose derivative sends, on the level of Lie algebras, $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \in \mathrm{sl}_2(\mathbb{C})$ to a nilpotent matrix of maximal rank in $\mathrm{gl}_n(\mathbb{C})$. These two extreme representations correspond to each other under a certain duality, generalising Alvis–Curtis duality for representations of $\mathrm{GL}_n(\mathbb{F}_q)$ (which itself generalises the duality between elementary symmetric polynomials e_n and homogeneous symmetric polynomials h_n in the context of representations of the symmetric group).
The examples of $\mathbf{1}$ and St illustrate Lusztig's idea of Jordan decomposition of representations: the subquotients of the parabolic induction of the supercuspidal representation μ should have the same semisimple part described by μ (and corresponding to the restriction of its L-parameter to W_K), and a varying unipotent part, captured by the $\mathrm{SL}_2(\mathbb{C})$ component of its L-parameter.

Following [Zelevinsky. Segments, 3.1], we work with *normalised* parabolic induction. A **segment** $[\rho_1, \rho_2]$ is a chain of irreducible supercuspidal representations $(\rho_1, \nu\rho_1, \cdots, \nu^k \rho_1 \cong \rho_2)$, where $\nu := |-| \circ \det$ (the twists by ν being necessary because of the normalisation). Two segments $[\rho_1, \rho_2]$ and $[\rho_3, \rho_4]$ are isomorphic if $\rho_1 \cong \rho_3$ and $\rho_2 \cong \rho_4$.
Given a segment $\Delta = [\rho, \nu^{k-1}\rho]$, with ρ an irreducible supercuspidal representation of $\mathrm{GL}_n(K)$, we obtain a representation $\rho \otimes \nu\rho \otimes \cdots \otimes \nu^{k-1}\rho$ of the standard Levi subgroup $\mathrm{GL}_m(K) \times \mathrm{GL}_m(K) \times \cdots \times \mathrm{GL}_m(K)$, which we can then parabolically induce to $\mathrm{GL}_n(K) = \mathrm{GL}_{mk}(K)$. Let us then denote by $\langle\Delta\rangle$ (respectively, $\langle\Delta\rangle^{\mathrm{t}}$) the unique irreducible subrepresentation (respectively, unique irreducible quotient) of this parabolically induced representation of $\mathrm{GL}_n(K)$ (uniqueness follows from [Zelevinsky. Segments, Proposition 2.1]). For instance, $\left\langle [\nu^{\frac{1-k}{2}}, \nu^{\frac{k-1}{2}}] \right\rangle$ is the trivial representation, and $\left\langle [\nu^{\frac{1-k}{2}}, \nu^{\frac{k-1}{2}}] \right\rangle^{\mathrm{t}}$ is the Steinberg representation.
The content of the Bernstein–Zelevinsky classification is that we can then obtain all smooth admissible representations of $\mathrm{GL}_n(K)$ by using these representations. More precisely, suppose we are given a collection of segments $\{\Delta_1, \ldots, \Delta_s\}$ fitting together to give representations $\langle\Delta_i\rangle$ of $\mathrm{GL}_{n_i}(K)$ with $\sum_{i=1}^s n_i = n$. Then we can *again* perform a parabolic induction to obtain a representation of $\mathrm{GL}_n(K)$, which we shall write

$$\langle\Delta_1\rangle \times \langle\Delta_2\rangle \times \cdots \times \langle\Delta_s\rangle.$$

Complications in the behaviour of this representation arise when some segments are **linked**: two segments Δ_{i_1}, Δ_{i_2} are said to be linked if $\Delta_{i_1} \not\subseteq \Delta_{i_2}$, $\Delta_{i_2} \not\subseteq \Delta_{i_1}$ and $\Delta_{i_1} \cup \Delta_{i_2}$ is also a segment. A special case is when Δ_{i_1} and Δ_{i_2} are **juxtaposed**, meaning that they are also disjoint. For instance, the segments $\Delta_{i_1} = [\rho_1, \rho_2]$ and $\Delta_{i_2} = [\rho_3, \rho_4]$ are juxtaposed if and only if $\rho_3 \cong \nu\rho_2$ (in which case we say Δ_{i_1} precedes Δ_{i_2}) or $\rho_1 \cong \nu\rho_4$. For instance, if Δ_1 precedes Δ_2, then instead of considering $\langle\Delta_1\rangle \times \langle\Delta_2\rangle$ it would have been better to consider $\langle\Delta_1 \cup \Delta_2\rangle$; this keeps the redundant information down to a minimum by ensuring that we pack all the complexity into a single step.
The basic result about the internal structure of these parabolic inductions is that $\langle\Delta_1\rangle \times \cdots \times \langle\Delta_s\rangle$ is irreducible if and only if none of the segments $\Delta_1, \ldots, \Delta_s$ are linked [Zelevinsky. Segments, Theorem 4.2]. The same also holds true for $\langle\Delta_1\rangle^{\mathrm{t}} \times \cdots \times \langle\Delta_s\rangle^{\mathrm{t}}$ [Zelevinsky. Segments, Theorem 9.7]. This demonstrates the usefulness of segments, as it provides the exact criterion for irreducibility of the parabolic induction of a tensor product of irreducible supercuspidal representations: there should not be a pair of representations ρ_1, ρ_2 with $\rho_2 \cong \nu\rho_1$. Segments then correspond to the cases where the parabolic induction is maximally far from being irreducible.
We can now state the Bernstein–Zelevinsky classification [Zelevinsky. Segments, Theorem 6.1]

Theorem 2.4 — Let $\Delta_1, \ldots, \Delta_s$ be segments as above. If Δ_i does not precede Δ_j for every $1 \leqslant i < j \leqslant s$, then $\langle\Delta_1\rangle \times \cdots \times \langle\Delta_s\rangle$ contains a unique irreducible subrepresentation, denoted $\langle\Delta_1, \ldots, \Delta_s\rangle$.
Every irreducible smooth admissible representation of $\mathrm{GL}_n(K)$ is isomorphic to a representation of the form $\langle\Delta_1, \ldots, \Delta_s\rangle$. Two such representations are isomorphic if and only if the corresponding sequences of segments are isomorphic, up to a rearrangement. ◇□

2.3.1 Relation with the local Langlands correspondence

The local Langlands conjectures lead us to expect a relation between irreducible smooth admissible representations of $G = \boldsymbol{G}(K)$ and representations of the absolute Galois group G_K of K. Working with complex representations of G_K, Lusztig's aforementioned principle of Jordan decomposition of characters makes us consider complex representations of the Weil–Deligne group $\mathrm{WD}_K = \mathrm{W}_K \times \mathrm{SL}_2(\mathbb{C})$, which are algebraic on the $\mathrm{SL}_2(\mathbb{C})$ component, and sending a Frobenius element of W_K to a semisimple element. Recall that W_K denotes the absolute Weil group of K.
In the case of $\boldsymbol{G} = \mathrm{GL}_n$, these representations exactly fit the bill:

Theorem 2.5 — There exists a unique set of bijections

$$\left\{\begin{array}{c}\text{Smooth admissible irreducible complex}\\ \text{representations of } \mathrm{GL}_n(K)\end{array}\right\} \xrightarrow{\mathrm{rec}_{K,n}} \left\{\begin{array}{c} n\text{-dimensional complex representations of } \mathrm{WD}_K \\ \text{which are algebraic and Frobenius-semisimple}\end{array}\right\}$$

which satisfies the following properties:

- $\mathrm{rec}_{K,1}$ recovers class field theory: $\mathrm{rec}_{K,1}(\chi) = \chi \circ \mathrm{Art}_K^{-1}$, where $\mathrm{Art}_K \colon K^\times \to \mathrm{W}_K^{\mathrm{ab}}$ is the Artin reciprocity map,
- $\mathrm{rec}_{K,n}$ commutes with twisting: $\mathrm{rec}_{K,n}(\pi \otimes (\chi \circ \det)) = \mathrm{rec}_{K,n}(\pi) \otimes \mathrm{rec}_{K,1}(\chi)$,
- $\mathrm{rec}_{K,n}$ is compatible with taking the central character: $(\det \circ \mathrm{rec}_{K,n})(\pi) = \mathrm{rec}_{K,1}(\omega_\pi)$, where ω_π is the central character of π,
- $\mathrm{rec}_{K,n}$ commutes with duality: $\mathrm{rec}_{K,n}(\pi^\vee) = \mathrm{rec}_{K,n}(\pi)^\vee$,
- $\mathrm{rec}_{K,n}$ matches up L-functions and ϵ-factors for pairs. ◇□

To illustrate the relation with the Berstein–Zelevinsky classification, we decompose representations of WD_K:

$$\rho = \bigoplus_{i=1}^{r} \rho_i \otimes \mathrm{Sp}(m_i),$$

where $\rho_1, \ldots, \rho_r$ are irreducible representations of W_K, and $\mathrm{Sp}(m_i) = \bigvee^{m_i - 1} \mathrm{Std}$ is the (unique up to isomorphism) irreducible m_i-dimensional algebraic complex representation of $\mathrm{SL}_2(\mathbb{C})$.
Writing $\pi_i = \mathrm{rec}_{K,n_i}^{-1}(\rho_i)$ (with $n_i = \dim(\rho_i)$), it then makes sense to set

$$\mathrm{St}_{m_i}(\pi_i) = \left\langle \left[\pi_i \nu^{\frac{1-m_i}{2}}, \pi_i \nu^{\frac{m_i-1}{2}} \right] \right\rangle,$$

$$\mathrm{rec}_{K,n}(\rho) = \langle \mathrm{St}_{m_1}(\pi_1) \times \cdots \times \mathrm{St}_{m_r}(\pi_r) \rangle .$$

In this way, we can reduce the problem of constructing a local Langlands correspondence for $\mathrm{GL}_n(K)$ to the supercuspidal case: it suffices to relate *supercuspidal* representations of $\mathrm{GL}_n(K)$ with *irreducible* representations of WD_K. Reducible representations of WD_K are then taken care of through the Bernstein–Zelevinsky classification. We go from irreducible to indecomposable representations of WD_K by the first induction step, with the $\mathrm{SL}_2(\mathbb{C})$ component taking care of the unipotent part as described previously (e.g. distinguishing between the trivial representation, with trivial L-parameter, and the Steinberg representation, whose L-parameter is trivial on W_K but is maximally nontrivial on $\mathrm{SL}_2(\mathbb{C})$). We then go from indecomposables to general (algebraic, Frobenius-semisimple) representations of WD_K with the second induction step of the Bernstein–Zelevinsky classification, corresponding to the direct sum of representations of WD_K.
As for semistability, we have the following:

Proposition 2.6 — The local Langlands reciprocity map $\mathrm{rec}_{K,n}$ matches up Iwahori-spherical representations of $\mathrm{GL}_n(K)$ with semistable representations of WD_K (i.e. representations of WD_K such that the restriction to W_K is unramified). ◇

Proof: Irreducible unramified complex representations of W_K are necessarily one dimensional, so that by the Bernstein–Zelevinsky classification the claim is equivalent to showing that the Iwahori-spherical representations are precisely the subquotients of unramified principal series representations. This was proved by Casselman (for general reductive groups) [Casselman. Unramified P.S. I, Proposition 2.6]. □

For more general reductive groups $\boldsymbol{G}$, we don't have the luxury of the Berstein–Zelevinsky classification. Instead, the **Deligne–Langlands conjecture** provides an explicit description of the local Langlands correspondence in the semistable case: on top of the semisimple element s (image of Frobenius) and unipotent element u (image of $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \mathrm{SL}_2(\mathbb{C})$), one has to add an additional ingredient: a representation of the component group of the simultaneous centraliser of s and u. See [Kazhdan–Lusztig], which tackles the case of a general split reductive group (with connected centre).

3 Iwahori–Hecke algebras

As we've seen in section 2.2, the representation theory of p-adic groups is controlled by underlying Hecke algebras. In the case of unramified representations, we would be working in the context of a reductive p-adic algebraic group $\boldsymbol{G}/K$ which is **unramified**, meaning that it has a model $\mathcal{G}/\mathcal{O}_K$ (a reductive group scheme over $\mathcal{O}_K$ with generic fibre $\boldsymbol{G}$).[70] From this integral model one constructs the *hyperspecial* maximal compact subgroup $U = \mathcal{G}(\mathcal{O}_K)$. The representation theory of unramified representations of G (i.e. U-spherical representations of G) is then controlled by the Hecke algebra $\mathcal{H}(G /\!\!/ U)$. The **Satake isomorphism** explicitly describes this Hecke algebra; in particular, it is commutative (as can be seen directly using Gelfand's lemma). This gives us the ability to classify all unramified representations with relative ease.

However, we are more interested in looking at Iwahori-spherical representations, as these are intimately connected to semistable p-adic Galois representations, as we saw in section 2.3.1. This means looking at the Hecke algebra $\mathcal{H}(G /\!\!/ I)$, where I is an Iwahori subgroup of G.

In the split case, the Satake isomorphism describes the spherical Hecke algebra $\mathcal{H}(G /\!\!/ U)$ as a q-deformation of the Weyl group W of $\boldsymbol{G}$; for the Iwahori–Hecke algebra $\mathcal{H}(G /\!\!/ I)$ we can proceed similarly, involving the *extended affine Weyl group* $\widetilde{W}$ of $\boldsymbol{G}$. This is the content of the **Iwahori–Matsumoto presentation**, which describes the algebra $\mathcal{H}(G /\!\!/ I)$ as spanned by elements T_w for $w \in \widetilde{W}$, subject to only the relations $T_{vw} = T_v T_w$ when $\ell(vw) = \ell(v) + \ell(w)$ (the braid relation) and $(T_s+1)(T_s-q) = 1$ for $s \in \widetilde{S}$, the subset of simple affine reflections of $\widetilde{W}$.

3.1 Double coset decompositions

We are interested in smooth admissible representations of G which admit non-zero I-fixed vectors, as these are closely related to semistable Galois representations as explained in section 1. Such representations correspond to representations of the Hecke algebra $\mathcal{H}(G /\!\!/ I)$ of compactly supported, bi-I-invariant functions on G. We want to compute the action of Hecke operators in $\mathcal{H}(G /\!\!/ I)$ on such representations; this information is needed, as explained in the introduction, to obtain the optimal integrality results later on when we will consider the action of Hecke operators on spaces of Hilbert modular forms.

The idea is that the double cosets in $I \backslash G / I$ are indexed by the extended affine Weyl group $\widetilde{W}$: this is the content of the **Iwahori–Bruhat decomposition** [Iwahori–Matsumoto, Theorem 2.16]. This allows us to reduce computations with double cosets in $I \backslash G / I$ to doing algebra in $\widetilde{W}$. One has [Lansky. Double cosets, Proposition 3.2]

Proposition 3.1 — For all $w \in \widetilde{W}$, $s \in \widetilde{S}$, we have:

- $IsIwI = IswI$ if $\ell(sw) > \ell(w)$,

[70] Equivalently, $\boldsymbol{G}$ is quasi-split (i.e. it contains a Borel subgroup defined over $\boldsymbol{K}$) and is split by an unramified extension of $\boldsymbol{K}$.

- $IsIwI = IswI \cup IwI$ if $\ell(sw) < \ell(w)$. ◇□

The second piece of information required to perform computations with I-double cosets is the ability to write down left coset representatives for double cosets [Lansky. Double cosets, Proposition 3.3]

Proposition 3.2 — If $s_\alpha \in S$ is the reflection associated to a simple (non-affine) root α, then

$$I s_\alpha I = \coprod_{[a] \in \mathcal{O}_K/(\varpi)} x_\alpha(a) \tilde{s}_\alpha I.$$

If $s_{\alpha,k} \in \widetilde{S}$ is the affine reflection associated to translation factor k and a highest root α of an irreducible root subsystem, then

$$I s_{\alpha,k} I = \coprod_{[a] \in \mathcal{O}_K/(\varpi)} x_{-\alpha}(\varpi a) \tilde{s}_{\alpha,k} I.$$

In both of these formulæ, $\tilde{s}$ denotes any chosen lift of s to the normaliser $N = \mathrm{N}_{\boldsymbol{G}}(\boldsymbol{T})(K)$, x_α denotes the root space homomorphism $x_\alpha \colon \mathbb{G}_{\mathrm{a}} \to \boldsymbol{G}$ attached to α, and the disjoint unions are over sets of representatives $a \in \mathcal{O}_K$ of $\mathcal{O}_K/(\varpi)$, where we take $0 \in \mathcal{O}_K$ to represent $0 \in \mathcal{O}_K/(\varpi)$. ◇□

Putting this together, we can compute double coset operators. For instance, for $\boldsymbol{G} = \mathrm{GL}_{n,K}$, we can immediately compute:

Proposition 3.3 — The double coset operator $[I t_{\varpi,i} I]$, where $t_{\varpi,i} = \mathrm{diag}(\overbrace{\varpi, \cdots, \varpi}^{i}, \overbrace{1, \cdots, 1}^{n-i})$, acts by the scalar $(q_\varpi)^{i(n-i)}$ on $(\mathrm{Ind}_B^G \mathbf{1}_B)^I$. ◇

Proof: It suffices to decompose the double cosets $[I t_{\varpi,i} I]$ into I-left cosets. According to the above two propositions, we find that these decompose into left cosets of upper triangular matrices, with $t_{\varpi,i}$ along the diagonal, and with entries $\mathbb{F}_{q_\varpi}$-equivalence-classes of elements placed at the coordinates (j,k) with $j \leqslant i$, $k \geqslant i+1$. For illustration, we have for instance, with $n = 5$ and $i = 3$:

$$I t_{\varpi,3} I = \coprod_{[a_{j,k}] \in \mathcal{O}_K/(\varpi)} \begin{pmatrix} p & 0 & 0 & a_{1,4} & a_{1,5} \\ 0 & p & 0 & a_{2,4} & a_{2,5} \\ 0 & 0 & p & a_{3,4} & a_{3,5} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} I$$

Each of the relevant matrices act trivially, and there are $(q_\varpi)^{i(n-i)}$ equivalence classes to consider. □

As a consequence, the action of the double coset operator $[I t_{\varpi,i} I]$ on the Steinberg representation, the unique irreducible quotient of $(\mathrm{Ind}_B^G \mathbf{1}_B)^I$, is also by the scalar $(q_\varpi)^{i(n-i)}$.

4 Hilbert modular forms

4.1 The context of Shimura varieties

Now that we have elucidated the representation-theoretic aspects of the theory of automorphic forms, we need to introduce the geometric objects which provide the required bridge to Galois representations.
In the case of classical modular forms, we saw in section 1 that the geometric objects of interest are the modular curves, where both modular forms and their associated Galois representations live. The key to generalisation is to reformulate the moduli problem these modular curves solve. Instead of considering modular curves as classifying elliptic curves (with extra structure), it is fruitful to consider that they classify (polarised) Hodge structures; this corresponds to the equivalence between complex abelian varieties and polarisable torsion-free $\mathbb{Z}$-Hodge structures of type $\{(-1,0),(0,-1)\}$. This equivalence is given by associating to A the Hodge structure $\mathrm{H}_1^{\mathrm{sing}}(A,\mathbb{Z})$, and to the Hodge structure V the complex torus $V_{\mathbb{C}}/(V^{-1,0}\oplus V_{\mathbb{Z}})$, which is algebraic thanks to the polarisation.

We can elaborate this description somewhat. Returning to the case of elliptic curves, we want to bring the algebraic group $\mathrm{GL}_{2,\mathbb{Q}}$ to the forefront. To do this, we have to perform a few reformulations. We start by considering free $\mathbb{Z}$-modules Λ of rank 2, equipped with a complex structure on $\Lambda_{\mathbb{R}} = \Lambda\otimes_{\mathbb{Z}}\mathbb{R}$ (i.e. a Hodge structure of type $\{(-1,0),(0,-1)\}$ on Λ). To bring in the action of GL_2, we fix the standard lattice $\mathbb{Z}^2\subset\mathbb{R}^2$, and obtain other lattices in $\mathbb{R}^2$ through the action of $\mathrm{GL}_2(\mathbb{Z})\backslash\mathrm{GL}_2(\mathbb{R})$. We first study the simpler problem of classifying elliptic curves up to isogeny, so we turn to $\mathbb{Q}$-lattices in $\mathbb{R}^2$, up to the action of $\mathrm{GL}_2(\mathbb{Q})\backslash\mathrm{GL}_2(\mathbb{R})$. We then use an adélic description of such $\mathbb{Q}$-lattices, by specifying $\mathbb{Q}_p$-lattices for all p simultaneously. We are thus interested in pairs consisting of an element of $\mathrm{GL}_2(\mathbb{A}_{\mathbb{Q}}^{(\infty)})$ and a complex structure on $\mathbb{R}^2$, up to the actions of $\mathrm{GL}_2(\mathbb{Q})$ and $\mathrm{GL}_2(\hat{\mathbb{Z}})$.
Now, a complex structure on $\mathbb{R}^2$ is given by the choice of a non-real line in $\mathbb{C}^2$, i.e. an element of $\mathfrak{X}=\mathbb{P}^1(\mathbb{C})\smallsetminus\mathbb{P}^1(\mathbb{R})$. This expresses the set of isogeny classes of elliptic curves over $\mathbb{C}$ as

$$\mathrm{Sh}_{\mathrm{GL}_2(\hat{\mathbb{Z}})}(\mathrm{GL}_{2,\mathbb{Q}},\mathfrak{X})_{\mathbb{C}}(\mathbb{C}) = \mathrm{GL}_2(\mathbb{Q})\backslash\mathfrak{X}\times\mathrm{GL}_2(\mathbb{A}_{\mathbb{Q}}^{(\infty)})/\mathrm{GL}_2(\hat{\mathbb{Z}}).$$

The situation in this case is slightly simplified, as all complex tori of dimension 1 are algebraic. Working with other groups $\boldsymbol{G}$, we want to classify *polarised* Hodge structures, instead of all Hodge structures.

This setup motivated Deligne to reformulate the work of Shimura, in order to define Shimura varieties for a linear algebraic group $\boldsymbol{G}$ as moduli spaces of polarisable variations of Hodge structure with Mumford–Tate group $\boldsymbol{G}$. The two main conditions are Griffiths transversality, and polarisability of the universal variation. These two conditions are enforced by the following definition, as provided in [Deligne. Shimura, 1.5] :

Definition 4.1 — A **Shimura datum** consists of a pair $(\boldsymbol{G},\mathfrak{X})$, where $\boldsymbol{G}$ is a reductive group over $\mathbb{Q}$ and $\mathfrak{X}\subseteq\mathrm{Hom}(\mathbb{S},\boldsymbol{G}_{\mathbb{R}})$ is a $\boldsymbol{G}(\mathbb{R})$-conjugacy class, subject to Deligne's axioms:

- For all $h\in\mathfrak{X}$, the induced Hodge structure on the adjoint representation of $\boldsymbol{G}_{\mathbb{R}}$ is of weight 0, and of type $\{(-1,1),(0,0),(1,-1)\}$.
- $\mathrm{ad}(h(\mathrm{i}))$ is a Cartan involution of $\boldsymbol{G}_{\mathbb{R}}^{\mathrm{ad}}$ for each $h\in\mathfrak{X}$. ◇

The Shimura variety attached to a Shimura datum can then be defined, generalising the case of modular curves. We saw above the double quotient

$$\mathrm{GL}_2(\mathbb{Q})\backslash\mathfrak{X}\times\mathrm{GL}_2(\mathbb{A}_{\mathbb{Q}}^{(\infty)})/U,$$

for $U = \mathrm{GL}_2(\hat{\mathbb{Z}})$; more generally we can consider this quotient for other compact open subgroups $U\subset\mathrm{GL}_2(\mathbb{A}_{\mathbb{Q}}^{(\infty)})$.
We can then associate to the Shimura datum $(\boldsymbol{G},\mathfrak{X})$ and the compact open subgroup $U\subseteq\boldsymbol{G}(\mathbb{A}_{\mathbb{Q}}^{(\infty)})$ the double quotient

$$\mathrm{Sh}_U(\boldsymbol{G},\mathfrak{X})_{\mathbb{C}}(\mathbb{C}) := \boldsymbol{G}(\mathbb{Q})\backslash\mathfrak{X}\times\boldsymbol{G}(\mathbb{A}_{\mathbb{Q}}^{(\infty)})/U.$$

The merit of Deligne's axioms is that they guarantee that this quotient has an algebraic structure (it is a stack, and for small enough U, a scheme). These turn out to have a model over a certain number field, the **reflex field** of the Shimura datum $(\boldsymbol{G}, \mathfrak{X})$. In general, these varieties *do not* have a moduli interpretation (say as a moduli spaces of abelian varieties).
However, we shall specialise to the case of Hilbert modular forms: as explained in the introduction, it is in that situation that we will be able to get a handle on the behaviour of Hecke operators on special representations, which will allow us to obtain a result mirroring Theorem I.14.4 in the automorphic context. This will allow us to give more explicit descriptions of the relevant geometric objects, which *do* have a moduli interpretation, which we shall review. After that is done, we shall move to the construction of **automorphic vector bundles**, which will provide the habitat for automorphic forms, generalising the line bundles $\omega(k)$ and $\omega_0(k)$ that we saw in section 1.

4.2 Moduli of Hilbert–Blumenthal abelian varieties

We now move on to the study of Hilbert modular forms, for a totally real number field F of degree d. To witness their geometric origins, we start by taking interest in Shimura varieties for the group $\boldsymbol{G} = \mathrm{Res}^F_{\mathbb{Q}} \mathrm{GL}_{2,F}$. However, the situation with Shimura varieties associated with $\boldsymbol{G}$ is complicated by the presence of a large determinant (the image of det being isomorphic to $\mathrm{Res}^F_{\mathbb{Q}} \mathbb{G}_{\mathrm{m},F}$ rather than $\mathbb{G}_{\mathrm{m},\mathbb{Q}}$). For this reason, it is convenient to introduce $\boldsymbol{G}^* = \boldsymbol{G} \times_{\mathrm{Res}^F_{\mathbb{Q}} \mathbb{G}_{\mathrm{m},F}, \det} \mathbb{G}_{\mathrm{m},\mathbb{Q}}$, whose attached Shimura varieties will have a simpler interpretation as moduli varieties than those attached to $\boldsymbol{G}$.
Let us now detail the relevant moduli problems.
A **Hilbert–Blumenthal abelian variety** (with respect to the totally real number field F) is an abelian variety of dimension $d = [F : \mathbb{Q}]$ equipped with a faithful action of $\mathcal{O}_F$. The yoga of PEL Shimura varieties then encourages us to think of Shimura varieties associated with $\boldsymbol{G}$ and $\boldsymbol{G}^*$ as classifying polarised Hilbert–Blumenthal abelian varieties equipped with a level structure. Recall that one possible definition of a polarisation of an abelian variety A is a morphism $\psi \colon A \to A^\vee$ satisfying the following two conditions:

- $\psi \colon A \to A^\vee$ is symmetric, meaning that the following diagram commutes:

$$\begin{tikzcd} A \arrow[dr, "\psi"] \arrow[dd, "\wr"'] & \\ & A^\vee \arrow[dl, "\psi^\vee"] \\ A^{\vee\vee} & \end{tikzcd}$$

- ψ is positive: passing to an algebraically closed base if necessary, ψ is attached to an *ample* line bundle $\mathfrak{L}$, through the formula $\psi(a) = (\mathrm{t}_a^* \mathfrak{L}) \otimes \mathfrak{L}^{-1} \in \mathrm{Pic}^0(A) = A^\vee$, where t_a denotes the morphism of translation by a.

Given a Hilbert–Blumenthal abelian variety A with real multiplication by $\mathcal{O}_F$, we can then consider the invertible $\mathcal{O}_F$-module $\mathrm{Hom}^{\mathrm{sym}}_{\mathcal{O}_F}(A, A^\vee)$ of symmetric homomorphisms $\psi \colon A \to A^\vee$ that commute with the action of $\mathcal{O}_F$. Inside $\mathrm{Hom}^{\mathrm{sym}}_{\mathcal{O}_F}(A, A^\vee)$, the polarisations form a cone $\mathrm{Hom}^{\mathrm{pol}}_{\mathcal{O}_F}(A, A^\vee)$. We are then interested in classifying Hilbert–Blumenthal abelian varieties with the pair $(\mathrm{Hom}^{\mathrm{sym}}_{\mathcal{O}_F}(A, A^\vee), \mathrm{Hom}^{\mathrm{pol}}_{\mathcal{O}_F}(A, A^\vee))$ running over the isomorphism classes $(\mathfrak{c}, \mathfrak{c}_+)$ of invertible $\mathcal{O}_F$-modules with a notion of positivity.[71]

Now, any abelian variety A over $\mathbb{C}$ is determined up to isomorphism by its homology group $\mathrm{H}_1(A, \mathbb{Z})$, as a polarisable $\mathbb{Z}$-Hodge structure. To give an action of $\mathcal{O}_F$ on A, it is equivalent to give $\mathrm{H}_1(A, \mathbb{Z})$ the structure of an $\mathcal{O}_F$-module.
Reformulating the problem, we can start with H, a projective $\mathcal{O}_F$-module of rank 2, and furnish $H \otimes_{\mathbb{Z}} \mathbb{R}$ with

[71] A notion of positivity on $\mathfrak{c}$ is an ordering on $\mathfrak{c} \otimes_{F,\tau} \mathbb{R}$ for each real embedding $\tau \colon F \hookrightarrow \mathbb{R}$. We then write $\mathfrak{c}_+$ for the subset of totally positive elements.

a complex structure, to then obtain a complex torus A, with $\mathcal{O}_F$ action, as the quotient $(H \otimes_{\mathbb{Z}} \mathbb{R})/H$. In this context a polarisation is simply a skew-symmetric morphism of Hodge structures $H \otimes_{\mathbb{Z}} H \to \mathbb{Z}(1)$; we get one for free thanks to the $\mathcal{O}_F$-action [Rapoport. Compactifications, 1.25].
The polarisation module $\mathrm{Hom}^{\mathrm{pol}}_{\mathcal{O}_F}(A, A^\vee)$ of A, constructed from the $\mathcal{O}_F$-module H, is then $\mathfrak{c} = (\bigwedge^2 H)^\vee(-1)$.

The interpretation of this construction via Shimura varieties is that Hilbert–Blumenthal abelian varieties, with given degree 1 homology H, are classified up to isogeny over $\mathbb{C}$ by the double quotient

$$\mathrm{GL}_{\mathcal{O}_F}(H)\backslash \mathrm{GL}_{\mathcal{O}_F\otimes_{\mathbb{Z}}\mathbb{R}}(H \otimes_{\mathbb{Z}} \mathbb{R})/(F \otimes_{\mathbb{Q}} \mathbb{C})^\times,$$

where $(F \otimes_{\mathbb{Q}} \mathbb{C})^\times$ is the stabiliser of any chosen complex structure on $H \otimes_{\mathbb{Z}} \mathbb{R}$, under the action by conjugation. If we take $H = \mathcal{O}_F \oplus \mathcal{O}_F$, then $\mathfrak{c} = \left(\bigwedge^2 H\right)^\vee(-1) = \mathcal{O}_F^\vee(-1) = \mathfrak{d}_F^{-1}(-1)$, so the polarisation module is the inverse different $\mathfrak{d}_F^{-1}$.
Working with this fixed polarisation module $\mathfrak{d}_F^{-1}$, but no longer fixing H, we need to make use of $\boldsymbol{G}^*$ to describe the relevant moduli spaces. Letting $\mathfrak{X}$ be the $\boldsymbol{G}^*(\mathbb{R})$-conjugacy class of

$$h\colon a + b\mathrm{i} \mapsto \left(\begin{pmatrix} a & -b \\ b & a \end{pmatrix}\right)_\tau,$$

it follows that complex Hilbert–Blumenthal abelian varieties, with given polarisation module $\mathfrak{d}_F^{-1}$, and level structure of level U, are classified up to isogeny by the Shimura variety for $(\boldsymbol{G}^*, \mathfrak{X})$, with complex points

$$\mathrm{Sh}_U(\boldsymbol{G}^*, \mathfrak{X})_{\mathbb{C}}(\mathbb{C}) = \boldsymbol{G}^*(\mathbb{Q})\backslash \mathfrak{X} \times \boldsymbol{G}^*(\mathbb{A}_{\mathbb{Q}}^{(\infty)})/U.$$

Over their reflex field $\mathbb{Q}$, the Shimura varieties $\mathrm{Sh}_U(\boldsymbol{G}^*, \mathfrak{X})$ are thus isomorphic to the moduli varieties $M(U, \mathfrak{d}_F^{-1})$. If we want to study the moduli varieties associated with more general polarisation modules $(\mathfrak{c}, \mathfrak{c}^+)$, we have to consider $\mathrm{GL}(H)$ for other rank 2 projective $\mathcal{O}_F$-modules than $H = \mathcal{O}_F \oplus \mathcal{O}_F$.[72]

The use of $\boldsymbol{G}^*$ instead of $\boldsymbol{G}$ in the above construction came about by our desire to preserve the polarisation module. Indeed, considering Shimura varieties as classifying Hodge structures with prescribed Mumford–Tate group, the determinant acts on polarisations. This means that if we want to work with $\boldsymbol{G}$, we have to account for the larger determinant; then, instead of classifying objects with polarisations (up to isogeny, i.e. action of $\mathbb{Q}^\times$), we would be considering objects where the polarisations are only given up to action by $F^\times$. This data (of a $\mathfrak{c}$-polarisation up to the action of $F^\times$) is called a **polarisation class**.
To classify Hilbert–Blumenthal abelian varieties with given polarisation class, we note that we are allowed to multiply the polarisation by elements of $\mathcal{O}_F^{\times,+}$, the group of totally positive units of F, without altering the pair $(\mathfrak{c}, \mathfrak{c}^+)$. That is, given $\varepsilon \in \mathcal{O}_F^{\times,+}$, we write

$$\varepsilon \cdot (\mathcal{A}, \iota, \psi, \lambda) = (\mathcal{A}, \iota, \varepsilon\psi, \lambda).$$

On the other hand, the endomorphism of $\mathcal{A}$ given by multiplication by $\varepsilon \in \mathcal{O}_F^{\times,+}$ (through ι) yields an isomorphism between $(\mathcal{A}, \iota, \psi)$ and $(\mathcal{A}, \iota, \varepsilon^2\psi)$. Let us pick an integer N such that U contains the principal congruence subgroup $U(N)$, and write $\mathcal{O}_{F,N}^\times = \left\{\varepsilon \in \mathcal{O}_F^\times \;\middle|\; \varepsilon \equiv 1 \pmod N\right\} \subseteq \mathcal{O}_F^\times$. The subgroup of squares $\mathcal{O}_{F,N}^{\times 2}$ in $\mathcal{O}_{F,N}^\times$ thus consists of units that act trivially on the level structure; in particular, the action of $\mathcal{O}_F^{\times,+}$ on polarisations is trivial when restricted to $\mathcal{O}_{F,N}^{\times 2}$. In other words, the action factors through the finite group $O_N := \mathcal{O}_F^{\times,+}/\mathcal{O}_{F,N}^{\times 2}$. We can then take the quotient under this action, yielding moduli stacks of Hilbert–Blumenthal abelian varieties with polarisation class $M^{\boldsymbol{G}}(U, \mathfrak{c}) := M(U, \mathfrak{c}) /\!\!/ O_N$. Taking $\mathfrak{c} = \mathfrak{d}_F^{-1}$, these quotients match up with Shimura varieties for the group $\boldsymbol{G}$. Even for small U, these do not have the structure of a scheme, and are instead only algebraic stacks; the word "varieties" is a misnomer.

[72] More precisely, for polarisation module $(\mathfrak{c}, \mathfrak{c}^+)$ we can take $\boldsymbol{H} = \mathcal{O}_F \oplus \mathfrak{c}^\vee$ and use $\mathrm{GL}^*(\boldsymbol{H}) := \{g \in \mathrm{GL}^*(\mathcal{O}_F \oplus \mathfrak{c}^\vee) \mid \det(g) \in \mathbb{Q} \subseteq \mathrm{GL}(\mathfrak{c}^\vee)\}$.

4.3 Toroidal compactifications of Hilbert moduli varieties

4.3.1 Minimal compactifications

We now turn to the task of compactifying the moduli spaces $M(U,\mathfrak{c})$, in order to deal with the behaviour at infinity. The obvious first step in this process is to compactify the Hermitian symmetric domains that occur in the Shimura-theoretic description of $M(U,\mathfrak{c})$. We prefer to work with connected Shimura varieties; this means working with the upper half-space $\mathcal{H}_F = \{(z_\tau)_\tau \in F \otimes_{\mathbb{Q}} \mathbb{C} \mid \forall \tau,\ \operatorname{im}(z_\tau) > 0\}$ (a connected component of the full Hermitian symmetric domain $\mathcal{X}$ used to define $M(U,\mathfrak{c})$), and the derived groups $\boldsymbol{G}^{\mathrm{der}}$ corresponding to $\mathrm{SL}(\mathcal{O}_F \oplus \mathfrak{c}^\vee)$ and $\boldsymbol{G}^{*,\mathrm{der}}$ corresponding to $\mathrm{SL}^*(\mathcal{O}_F \oplus \mathfrak{c}^\vee)$.

The rational boundary component for $\mathcal{H}_F$ is given by $\mathbb{P}^1(F) \hookrightarrow \prod_{\tau: F\to\mathbb{R}} \mathbb{P}^1(\mathbb{Q})$. Analytically, we can then compactify $\mathcal{H}_F$ by giving $\mathcal{H}_F^* := \mathcal{H}_F \coprod \mathbb{P}^1(F)$ the **Satake topology**, which is determined by giving a fundamental system of (punctured) neighbourhoods of the cusp ∞:

$$W(\infty, r) := \left\{ z \in \mathcal{H}_F \;\middle|\; \prod_{\tau: F\hookrightarrow\mathbb{R}} \operatorname{im}(z_\tau) > r \right\},$$

and using the action of $\boldsymbol{G}^*(\mathbb{Q})$ to obtain neighbourhoods of the other cusps. This indeed yields an algebraic object, called the **minimal compactification** of $M(U,\mathfrak{c})$.[73]

One problem arises however: the resulting complex analytic spaces are singular for $F \neq \mathbb{Q}$ (i.e. $d > 1$). To see this, let's begin by working with the cusp ∞. We have:

$$\operatorname{Stab}_U(\infty) \overset{[74]}{=} \operatorname{Stab}_U(W(\infty,r)) = \boldsymbol{B}(\mathbb{Q}) \cap U = \left\{ \begin{pmatrix} \varepsilon & c \\ 0 & \varepsilon^{-1} \end{pmatrix} \;\middle|\; \varepsilon \in \mathcal{O}_F^\times, c \in \mathfrak{c}^\vee \right\} \cap U \subset \boldsymbol{G}^{*,\mathrm{der}}(\mathbb{Q})$$

for large enough r, where $\boldsymbol{B}$ denotes the standard (upper triangular) Borel subgroup of $\boldsymbol{G}^{*,\mathrm{der}}$.

Writing $B_U := \boldsymbol{B}(\mathbb{Q}) \cap U$, the Levi decomposition $\boldsymbol{B} = \boldsymbol{T}\boldsymbol{V}$ (with $\boldsymbol{T}$ the maximal torus associated with $\boldsymbol{B}$, and $\boldsymbol{V}$ the unipotent radical of $\boldsymbol{B}$) means that we can consider taking the quotient $B_U \backslash \mathcal{H}_F$ in two successive steps: first quotient by the unipotent part $V_U := \boldsymbol{V}(\mathbb{Q}) \cap U = \left\{ \begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix} \;\middle|\; c \in \mathfrak{c}^\vee \right\}$, and then by the toric part $T_U := \boldsymbol{T}(\mathbb{Q}) \cap U = \left\{ \begin{pmatrix} \varepsilon & 0 \\ 0 & \varepsilon^{-1} \end{pmatrix} \;\middle|\; \varepsilon \in \mathcal{O}_F^\times \right\} \cap U$. Note that T_U is isomorphic to a subgroup of finite index of $\mathcal{O}_F^\times$, but with the action of $\varepsilon \in \mathcal{O}_F^\times$ given by multiplication by ε^2.

Embedding $\mathcal{H} \hookrightarrow \mathbb{C}$, the situation thus looks like we are quotienting $F \otimes_{\mathbb{Q}} \mathbb{C}$ first by $V_U \cong \mathfrak{c}^\vee$ acting by translation via the formula

$$c \cdot (z_1, \ldots, z_d) = (z_1 + \tau_1(c), \ldots, z_d + \tau_d(c)),$$

and second by T_U.

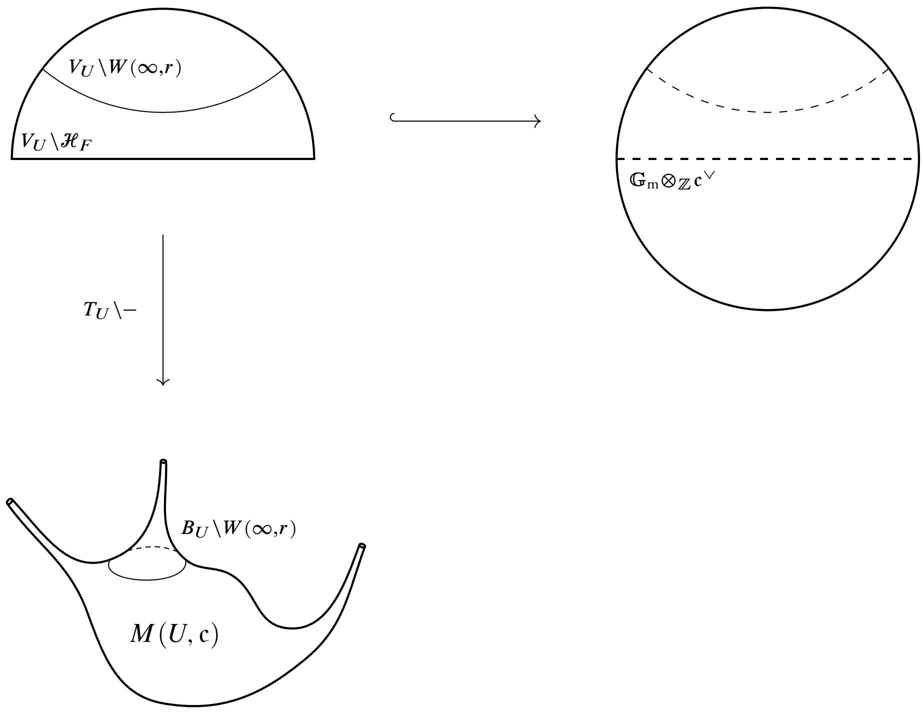
Through the exponential exact sequence

$$0 \longrightarrow \mathfrak{c}^\vee \longrightarrow F \otimes_{\mathbb{Q}} \mathbb{C} \longrightarrow \mathbb{G}_{\mathrm{m}} \otimes_{\mathbb{Z}} \mathfrak{c}^\vee \longrightarrow 0,$$

the neighbourhoods $V_U \backslash W(\infty, r)$ thus look like analytic neighbourhoods of $\mathbb{G}_{\mathrm{m}} \otimes_{\mathbb{Z}} \mathfrak{c}^\vee$ around 0, and further quotienting by T_U then yields the fundamental neighbourhoods $B_U \backslash W(\infty, r)$ of the cusp ∞ in $M(U,\mathfrak{c})$. We can picture the situation for $[F : \mathbb{Q}] = 2$ as depicted in [A–M–R–T, §I.5]

[73] To prove the algebraicity of this object, it is easiest to first construct the toroidal compactifications (as we are about to do), and then blow down to a point the various components that were added in the process.

[74] This first equality is valid for $\boldsymbol{G}^*$ but not for $\boldsymbol{G}$, in which case $\operatorname{Stab}_U(\infty)$ is larger, with corresponding matrices of the form $\begin{pmatrix} u\varepsilon & c \\ 0 & \varepsilon^{-1} \end{pmatrix}$ with $u \in \mathcal{O}_F^{\times,+}$. The rest remains unchanged.



Note that, around the cusps, $M(U,\mathfrak{c})$ is analytically akin to the quotient of $\mathbb{C}^d$ by $(\zeta_1,\ldots,\zeta_d)$, acting by componentwise multiplication, for some roots of unity $\zeta_1,\ldots,\zeta_d$ of fixed order m. The image of $(0,\ldots,0)$ in this quotient is a singular point for $d>1$. Indeed, as $\mathbb{C}^d \smallsetminus (0,\ldots,0)$ is simply connected for $d>1$, the relative homology in degree 1 of the quotient with respect to the image of $(0,\ldots,0)$ is a finite cyclic group of order m.

4.3.2 Toric embeddings

The point of introducing $\mathbb{G}_m \otimes_{\mathbb{Z}} \mathfrak{c}^\vee$ is that we can make use of it to resolve the singularities that were introduced in quotienting by T_U. As we saw, $V_U \backslash W(\infty,r)$ embeds into $\mathbb{G}_m \otimes_{\mathbb{Z}} \mathfrak{c}^\vee$ as a neighbourhood of 0, and the action of T_U on $W(\infty,r)$ extends to $\mathbb{G}_m \otimes_{\mathbb{Z}} \mathfrak{c}^\vee$. We can then compactify $\mathbb{G}_m \otimes_{\mathbb{Z}} \mathfrak{c}^\vee$ by adding additional components, in a way that allows the action of T_U to extend. To carry out this process, we proceed as follows.
The torus $\boldsymbol{T}_\mathfrak{c} := \mathbb{G}_{m,\mathbb{Q}} \otimes_{\mathbb{Z}} \mathfrak{c}^\vee$ is uniquely determined by its character group $\mathrm{X}^*(\boldsymbol{T}_\mathfrak{c}) = \mathfrak{c}$ (as a Galois module with trivial Galois action),[75] and we can write

$$\boldsymbol{T}_\mathfrak{c} = \mathrm{Spec}\left(\mathbb{Q}\left[\mathrm{X}^*(\boldsymbol{T}_\mathfrak{c})\right]\right).$$

A natural way of (partially) compactifying $\boldsymbol{T}_\mathfrak{c}$ consists of inverting less elements in the above description; for instance we have $\mathbb{G}_m \subset \mathbb{A}^1$ given by $\mathrm{Spec}\left(\mathbb{Z}[t,t^{-1}]\right) \hookrightarrow \mathrm{Spec}\left(\mathbb{Z}[t]\right)$. Going even further, we can glue together these partial compactifications; for instance we can glue $\mathrm{Spec}\left(\mathbb{Z}[t]\right)$ and $\mathrm{Spec}\left(\mathbb{Z}[t^{-1}]\right)$ along $\mathrm{Spec}\left(\mathbb{Z}[t,t^{-1}]\right)$ to obtain the compactification $\mathbb{G}_m \subset \mathbb{P}^1$. Coming back to $\boldsymbol{T}_\mathfrak{c}$, we can consider multiplicative submonoids S of $\mathrm{X}^*(\boldsymbol{T}_\mathfrak{c})$, and study the partial compactifications $\mathrm{Spec}\left(\mathbb{Q}[S]\right)$.
In general, given a torus $\boldsymbol{T}$ over a field K, we define **toric embeddings** of $\boldsymbol{T}$ as open embeddings $\boldsymbol{T} \hookrightarrow Z$, such that Z is equipped with an action of $\boldsymbol{T}$ extending the action of $\boldsymbol{T}$ on itself by multiplication. We want to construct such embeddings using the above process. We begin with a reformulation in terms of combinatorial data.
Let $X^* = \mathrm{X}^*(\boldsymbol{T})$ be the character group of $\boldsymbol{T}$, and $X_* = \mathrm{X}_*(\boldsymbol{T})$ its cocharacter group. Individual affine toric embeddings $\boldsymbol{T}$ are then best described by convex cones $\sigma \subseteq (X_*)_\mathbb{R}$ (instead of multiplicative submonoids of X^*). Given a convex cone $\sigma \subseteq (X_*)_\mathbb{R}$, we consider the dual convex cone $\breve{\sigma} = \{x \in X^*_\mathbb{R} \mid \forall w \in \sigma, \langle x,w\rangle \geqslant 0\}$, and define

$$\overline{\boldsymbol{T}}_\sigma := \mathrm{Spec}\left(K[\breve{\sigma} \cap X^*]\right).$$

[75] Note that this makes $\mathbb{G}_m \otimes_{\mathbb{Z}} \mathfrak{c}^\vee$ into an $\mathcal{O}_F$-module scheme (“schéma en $\mathcal{O}_F$-modules”).

When we have two convex cones $\sigma_1, \sigma_2 \subseteq (X_*)_{\mathbb{R}}$ intersecting in a common face, we can glue $\overline{\boldsymbol{T}}_{\sigma_1}$ and $\overline{\boldsymbol{T}}_{\sigma_2}$ along $\overline{\boldsymbol{T}}_{\sigma_1 \cap \sigma_2}$.
Now, we can't just use any cones σ. For instance, to have a morphism $\boldsymbol{T} \hookrightarrow \overline{\boldsymbol{T}}_{\sigma}$ we require that $\breve{\sigma} \cap X^*$ generates X^* as a group. Additionally, $\overline{\boldsymbol{T}}_{\sigma}$ is a normal variety if and only if $\breve{\sigma} \cap X^*$ is saturated in X^* [K–K–M–S-D, I, Lemma 1]. These conditions are formulated in terms of $\breve{\sigma}$, but it is more convenient to transfer them to σ. To do so, we need to introduce a few notions:

- σ is **polyhedral** if it is the intersection of finitely many half-spaces,
- σ is moreover **rational** if these half-spaces can be taken to be rational,
- σ is **salient** if it does not contain any nonzero vector subspace of $(X_*)_{\mathbb{R}}$,
- σ is **smooth** if $\sigma \cap X_*$ is spanned by a subset of a ($\mathbb{Z}$-)basis of X_*.

We need σ to be a polyhedral rational convex cone to ensure that $\breve{\sigma} \cap X^*$ is a finitely generated monoid, whereas σ being salient ensures that $\overline{\boldsymbol{T}}_{\sigma}$ is of full dimension, thus containing $\boldsymbol{T}$.
The upshot of working with cones in $(X_*)_{\mathbb{R}}$ rather than multiplicative submonoids of X^* is the following elegant result [K–K–M–S-D, I, Theorem 1', Theorem 4] :

Proposition 4.2 — The application

$$\sigma \mapsto \overline{\boldsymbol{T}}_{\sigma} := \operatorname{Spec}\left(K[\breve{\sigma} \cap X^*]\right)$$

restricts to a morphism of categories between the category of salient polyhedral rational convex cones in $(X_*)_{\mathbb{R}}$ and the category of normal affine toric embeddings of $\boldsymbol{T}$, which is an equivalence of categories if K is algebraically closed. Additionally, $\overline{\boldsymbol{T}}_{\sigma}$ is smooth if and only if σ is smooth. ◇□

The morphisms in the previous proposition are as follows: a morphism of convex cones is simply a linear embedding, whereas morphisms between normal affine toric embeddings are required by definition to be equivariant for the $\boldsymbol{T}$-action.
The idea is that $\overline{\boldsymbol{T}}_{\sigma}$ partially compactifies $\boldsymbol{T}$ by adding limit points $\lim_{t \to 0} \mu(t)$ for cocharacters $\mu \in \sigma \cap X_*$ [K–K–M–S-D, I, Theorem 1'].

The next step is to glue together these affine partial compactifications. For this, we need to impose suitable conditions on collections of cones that allow for the gluing to take place. We say a (not necessarily finite) collection $\Sigma = \{\sigma \subseteq (X_*)_{\mathbb{R}}\}$ of convex cones is a **fan** if the following two conditions hold:

- all faces of cones in Σ are also in Σ,
- the intersection of any two cones in Σ is a face of those two cones.

Then, given a fan of salient polyhedral rational convex cones Σ, we can glue together the various $\overline{\boldsymbol{T}}_{\sigma}$ for $\sigma \in \Sigma$ to obtain $\overline{\boldsymbol{T}}_{\Sigma}$. Working over an algebraically closed field, these are in fact all the toric embeddings [K–K–M–S-D, I, Theorem 6, Theorem 8] :

Proposition 4.3 — The application

$$\Sigma \mapsto \overline{\boldsymbol{T}}_{\Sigma}$$

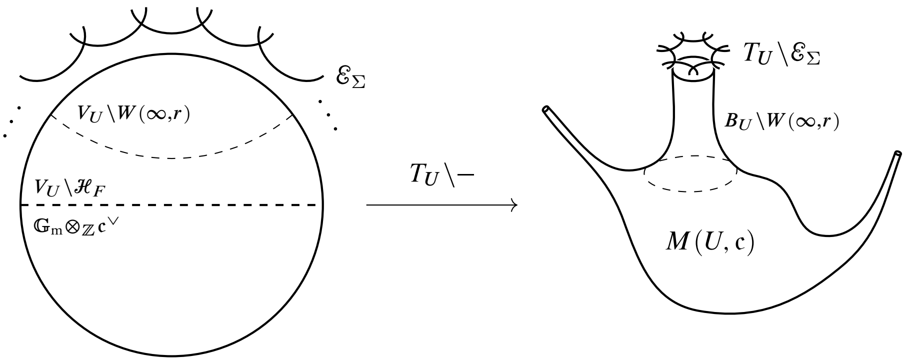
defines a functor between the category of fans of salient polyhedral rational convex cones in $X^*_{\mathbb{R}}$ and the category of normal toric embeddings of $\boldsymbol{T}$, which is an equivalence if K is algebraically closed.
Additionally, $\overline{\boldsymbol{T}}_{\Sigma}$ is of finite type over K if and only if Σ is finite; $\overline{\boldsymbol{T}}_{\Sigma}$ is then proper over K if and only if $\bigcup_{\sigma \in \Sigma} \sigma = (X_*)_{\mathbb{R}}$ (in which case we call Σ **complete**), and smooth over K if and only if all $\sigma \in \Sigma$ are smooth (in which case we call Σ **smooth**). ◇□

In this proposition, a morphism of fans $\Sigma_1 \to \Sigma_2$ consists of a linear map $f\colon (X_*)_{\mathbb{R}} \to (X_*)_{\mathbb{R}}$ such that for all $\sigma_1 \in \Sigma_1$, there is a $\sigma_2 \in \Sigma_2$ with $f(\sigma_1) \subseteq \sigma_2$.

4.3.3 Toroidal compactifications

Finally, we can make use of these results to compactify the Hilbert modular varieties. According to principles of log geometry, we expect a compactification with a divisor with normal crossings at infinity; this will allow Hodge theoretic results to extend to the compactification by using the methods of section I.9.6.
To bring into play the theory of toric embeddings, we make use of the embedding $V_U\backslash\mathcal{H}_F \hookrightarrow \mathbb{G}_m \otimes_{\mathbb{Z}} \mathfrak{c}^\vee$, where we recall that $V_U = \begin{pmatrix} 1 & \mathfrak{c}^\vee \\ 0 & 1 \end{pmatrix} \subset \mathrm{SL}^*(\mathcal{O}_F \oplus \mathfrak{c}^\vee)$, acting by translation on $\mathcal{H}_F$. The exponential exact sequence also gave rise to the description of $\boldsymbol{T}_{\mathfrak{c}} = \mathbb{G}_m \otimes_{\mathbb{Z}} \mathfrak{c}^\vee$ as $\mathfrak{c}^\vee\backslash(F \otimes_{\mathbb{Q}} \mathbb{C})$. Now, $\mathfrak{c}^\vee$ is a lattice in $F \otimes_{\mathbb{Q}} \mathbb{R}$, so that the torus $\boldsymbol{T}_{\mathfrak{c}}$ can be written as a direct product $\boldsymbol{T}_{\mathfrak{c}} = (\mathfrak{c}^\vee\backslash F \otimes_{\mathbb{Q}} \mathbb{R}) \times (F \otimes \mathrm{i}\mathbb{R})$. Write $\boldsymbol{T}_{\mathfrak{c},c} := \mathfrak{c}^\vee\backslash F \otimes_{\mathbb{Q}} \mathbb{R}$ (the maximal compact subtorus of $\boldsymbol{T}_{\mathfrak{c}}$), and let ord: $\boldsymbol{T}_{\mathfrak{c}} \to F \otimes_{\mathbb{Q}} \mathbb{R}$ be the "imaginary part" morphism associated with the projection onto the second factor in the previous direct product. The quotient $\mathfrak{c}^\vee\backslash\mathcal{H}_F$ is then $\mathrm{ord}^{-1}((F \otimes_{\mathbb{Q}} \mathbb{R})_+)$, whereas the fundamental neighbourhoods of the cusp at infinity are given by $\mathfrak{c}^\vee\backslash W(\infty, r) = \mathrm{ord}^{-1}\{(x_\tau)_\tau \in (F \otimes_{\mathbb{Q}} \mathbb{R})_+ \mid \prod_\tau x_\tau > r\}$. This means that if we want to use the theory of toric embeddings to compactify $M(U, \mathfrak{c})$, we need only add components by considering cones in the positive orthant $(F \otimes_{\mathbb{Q}} \mathbb{R})_+$ of $F \otimes_{\mathbb{Q}} \mathbb{R}$ (strictly speaking, we need to also add the origin back, so that we would be considering cones in $(F \otimes_{\mathbb{Q}} \mathbb{R})_+ \cup \{0\}$).
Let us now do so: consider a fan Σ of salient polyhedral rational convex cones in the positive orthant $(F \otimes_{\mathbb{Q}} \mathbb{R})_+$ of $F \otimes_{\mathbb{Q}} \mathbb{R}$, and complete within that orthant (so that $\bigcup_{\sigma\in\Sigma} \sigma = \{0\} \cup (F \otimes_{\mathbb{Q}} \mathbb{R})_+$). There will necessarily be infinitely many cones in Σ, as the facets of the cones are not allowed to be contained in any of the coordinate hyperplanes. On top of this, we recall that we have yet to quotient by T_U; so to get a quotient of finite type we require the fan to be preserved under the action of T_U, and such that there are only finitely many cones modulo this action.

We can then picture the compactification attached to such a complete fan Σ of $(F \otimes \mathbb{R})_+$, for $[F : \mathbb{Q}] = 2$, like in [A–M–R–T, I, §5] :



We then need to perform this procedure at all cusps, in a way that is compatible with the action of $\boldsymbol{G}^*(\mathbb{Q})$. This is always possible [Chai. Compactifications, §3.3].
The resulting compactifications are the **toroidal compactifications** $\overline{M}_\Sigma(U, \mathfrak{c})$ of $M(U, \mathfrak{c})$.

4.4 Automorphic forms for G^*

Armed with these compactifications, we can now define automorphic forms for the group $\boldsymbol{G}^*$. To do so, we choose automorphic vector bundles over $M(U, \mathfrak{c})$. We start with the vector bundle $\omega = \pi_*\omega_{\mathcal{A}}$, where $\pi: \mathcal{A} \to M(U, \mathfrak{c})$ is the universal Hilbert–Blumenthal abelian scheme over $M(U, \mathfrak{c})$. This is a locally free $\mathbb{O}_{M(U,\mathfrak{c})} \otimes_{\mathbb{Z}} \mathcal{O}_F$-module of rank 1. If we pick a field extension E/F containing the normal closure of F, this vector bundle thus decomposes over $M(U, \mathfrak{c})_E$ as $\omega = \bigoplus_{\tau: F\hookrightarrow E} \omega_\tau$.

Now, weights for $\boldsymbol{G}^*$ are given by d-tuples of integers $(\kappa_\tau)_\tau$, labelled by the embeddings $\tau: F \hookrightarrow \mathbb{R}$. Given a weight $\kappa = (\kappa_\tau)_\tau$, we define the associated automorphic vector bundle $\omega(\kappa)$ on $M(U, \mathfrak{c})_E$ by

$$\omega(\kappa) := \bigotimes_{\tau : F \hookrightarrow E} \omega_\tau^{\otimes \kappa_\tau}.$$

We are only interested in the **algebraic** weights; the automorphic forms of arithmetic interest (from the perspective of Galois representations) are those with algebraic weight. In this case, a weight $(\kappa_\tau)_\tau$ is algebraic if all κ_τ are positive integers of the same parity, and is moreover **regular** if all the weights are at least 2.

The next step is to extend these automorphic vector bundles to the toroidal compactifications $\overline{M}_\Sigma(U, \mathfrak{c})$. We omit the details for the moment, as we will cover this procedure in greater detail for integral models of $M(U, \mathfrak{c})$ in section 4.8. Suffices to say that we can naturally extend ω to the toroidal compactifications, and that it remains locally free as an $\mathcal{O}_F \otimes_{\mathbb{Z}} \mathcal{O}_{\overline{M}_\Sigma(U,\mathfrak{c})}$-module, allowing us to also extend the bundles $\omega(\kappa)$. We will use the same notation to denote the extension of ω and $\omega(\kappa)$ to the toroidal compactifications.
We can then finally define the space of automorphic forms of level U, polarisation module $\mathfrak{c}$ and weight κ. This is

$$\mathrm{M}_\kappa^{\boldsymbol{G}^*}(U, \mathfrak{c}; E) := \mathrm{H}^0(\overline{M}_\Sigma(U, \mathfrak{c})_E, \omega(\kappa)).$$

We will see in section 4.8 that this definition is in fact independent of the choice of Σ, which justifies the absence of Σ in the left hand side of the above definition.

4.5 Automorphic forms for G

However, we are more interested in automorphic forms for the group $\boldsymbol{G}$, as these are most closely related to Hilbert modular forms.
As we saw in section 4.2, to relate these to moduli spaces of Hilbert–Blumenthal abelian varieties, we must weaken "polarisation" to "polarisation class": instead of considering polarisations up to action by $\mathbb{Q}^\times$, we consider them up to action by $F^\times$. The downside is that these moduli spaces are never varieties (unless we resign to only considering them as coarse moduli spaces).
From the perspective of automorphic bundles on Shimura varieties attached to $\boldsymbol{G}$, weights are now given by pairs, consisting of a d-tuple $\nu = (\nu_\tau)_\tau$ of integers, labelled by embeddings $\tau : F \hookrightarrow \mathbb{R}$, together with an additional integer w. In this case, algebraicity is the requirement that w be of the same parity as the ν_τ.
Again choosing E/F containing the normal closure of F, we can define the automorphic vector bundle attached to such a weight (ν, w), at level U. To do this, we need the following two vector bundles on $M^{\boldsymbol{G}}(U, \mathfrak{c})$:

- ω, given by the same definition as in the case of $\boldsymbol{G}^*$: $\omega := \pi_* \omega_{\mathcal{A}}$, where $\mathcal{A}$ is the universal abelian scheme over $M^{\boldsymbol{G}}(U, \mathfrak{c})$.
- δ, given by $\delta := \bigwedge^2_{\mathcal{O}_{M^{\boldsymbol{G}}(U,\mathfrak{c})} \otimes_{\mathbb{Z}} \mathcal{O}_F} \mathcal{H}^1_{\mathrm{dR}}(\mathcal{A}/M^{\boldsymbol{G}}(U, \mathfrak{c}))$, with $\mathcal{A}$ as above.

In a similar manner to before, these vector bundles again decompose (once we base change to E) as a direct sum of line bundles ω_τ, δ_τ.
The vector bundle δ also has a simple description: the Hodge filtration of the universal Hilbert–Blumenthal abelian scheme $\mathcal{A}$ is given by the short exact sequence

$$0 \longrightarrow \omega \longrightarrow \mathcal{H}^1_{\mathrm{dR}}\left(\mathcal{A}/M^{\boldsymbol{G}}(U, \mathfrak{c})\right) \longrightarrow \omega^\vee \otimes_{\mathbb{Z}} \mathfrak{c}\mathfrak{d}_F^{-1} \longrightarrow 0.$$

Because ω is an invertible $\mathcal{O}_{M^{\boldsymbol{G}}(U,\mathfrak{c})} \otimes_{\mathbb{Z}} \mathcal{O}_F$-module, we have that

$$\delta = \bigwedge^2_{\mathcal{O}_{M^{\boldsymbol{G}}(U,\mathfrak{c})} \otimes_{\mathbb{Z}} \mathcal{O}_F} \mathcal{H}^1_{\mathrm{dR}}\left(\mathcal{A}/M^{\boldsymbol{G}}(U, \mathfrak{c})\right) \cong \mathfrak{c}\mathfrak{d}_F^{-1} \otimes_{\mathbb{Z}} \mathcal{O}_{M^{\boldsymbol{G}}(U,\mathfrak{c})}.$$

This shows that δ is free (not just locally free), given by $\mathcal{O}_{M^{\boldsymbol{G}}(U,\mathfrak{c})} \otimes_{\mathbb{Z}} \mathfrak{c}\mathfrak{d}_F^{-1}$.
Given a d-tuple of integers $(k_\tau)_\tau$ indexed by embeddings $F \overset{\tau}{\hookrightarrow} \mathbb{R}$, we can then define the line bundles $\omega(k)$ and

$\delta(k)$ on $M^{\boldsymbol{G}}(U, \mathfrak{c})_E$ as

$$\omega(k) := \bigotimes_{\tau : F \hookrightarrow E} \omega_\tau^{\otimes k_\tau}, \qquad \delta(k) := \bigotimes_{\tau : F \hookrightarrow E} \delta_\tau^{\otimes k_\tau}.$$

Finally, for weight (ν, w), let $l_\tau = \frac{w - \nu_\tau}{2}$, $l = (l_\tau)_\tau$. Then we define $\omega(\nu, w) := \omega(\nu) \otimes_{\mathcal{O}_{M^{\boldsymbol{G}}(U,\mathfrak{c})_E}} \delta(l)$.

These vector bundles also extend to the toroidal compactifications (see section 4.8). This allows us to define the space of automorphic forms for $\boldsymbol{G}$ of level U, polarisation module $\mathfrak{c}$ and weight (ν, w) as

$$\mathrm{M}^{\boldsymbol{G}}_{(\nu,w)}(U, \mathfrak{c}; E) := \mathrm{H}^0\left(\overline{M}^{\boldsymbol{G}}_\Sigma(U, \mathfrak{c})_E, \omega(\nu, w)\right).$$

As in the previous section, this space also does not depend on the choice of Σ; see section 4.8.

To connect this situation with that of $\boldsymbol{G}^*$, recall the setup from section 4.2: assuming $U(N) \subseteq U$, the action of $\mathcal{O}_F^{\times,+}$ on $M(U, \mathfrak{c})$ factors through the finite group $O_N = \mathcal{O}_F^{\times,+}/\mathcal{O}_{F,N}^{\times 2}$, and the quotient stack $M(U, \mathfrak{c}) /\!\!/ O_N$ is none other than $M^{\boldsymbol{G}}(U, \mathfrak{c})$. To carry this through to the automorphic vector bundles, we thus need to extend this action of $\mathcal{O}_F^{\times,+}$ to $\omega(\kappa)$.
Following [Andreatta–Iovita–Pilloni, §1], we begin by remarking that at the level of sections of $\omega(\kappa)$, the automorphism of $\mathcal{A}$ given by multiplication by $\varepsilon \in \mathcal{O}_{F,N}^\times$ yields equalities

$$f(\mathcal{A}, \iota, \psi, \lambda, \omega) = f(\mathcal{A}, \iota, \varepsilon^2\psi, \lambda, \varepsilon\omega) = \kappa(\varepsilon)^{-1} f(\mathcal{A}, \iota, \varepsilon^2\psi, \lambda, \omega).$$

Thus, if we want to define an action of $\mathcal{O}_F^{\times,+}$ on the automorphic vector bundles $\omega(\kappa)$ which is trivial on $\mathcal{O}_{F,N}^{\times 2}$, we need to have the equality

$$(\varepsilon^2 \cdot f)(\mathcal{A}, \iota, \psi, \lambda, \omega) = \kappa(\varepsilon) f(\mathcal{A}, \iota, \varepsilon^{-2}\psi, \lambda, \omega).$$

To achieve this, we choose ν and w such that $\kappa_\tau = 2\nu_\tau + w$ for all τ, and define the action by

$$(\varepsilon \cdot f)(\mathcal{A}, \iota, \psi, \lambda, \omega) = \nu(\varepsilon) f(\mathcal{A}, \iota, \varepsilon^{-1}\psi, \lambda, \omega).$$

This indeed works, as then $\nu(\varepsilon)^2 = \kappa(\varepsilon)$ because $\kappa = 2\nu + w$ and ε is a unit.
Once this action is fixed, we can then take the quotient; this recovers the automorphic vector bundle $\omega(\nu, w)$ on $\overline{M}^{\boldsymbol{G}}(U, \mathfrak{c})_E$.

We can go one step further than quotienting by the action of $\mathcal{O}_F^{\times,+}$, however, by using the action of $F^{\times,+}$. Now, elements of $F^{\times,+}$ can permute around the polarisation modules $\mathfrak{c}$. Given $x \in F^{\times,+}$, define an isomorphism

$$L_x \colon \mathrm{M}^{\boldsymbol{G}}_{(\nu,w)}(U, \mathfrak{c}; E) \longrightarrow \mathrm{M}^{\boldsymbol{G}}_{(\nu,w)}(U, x\mathfrak{c}; E)$$

by the same formula as above:

$$(L_x f)(\mathcal{A}, \iota, \psi, \lambda, \omega) = \nu(x) f(\mathcal{A}, \iota, x^{-1}\psi, \lambda, \omega).$$

We can then remove the polarisations by defining:

$$\mathrm{M}^{\boldsymbol{G}}(U; E) := \bigoplus_{(\mathfrak{c}, \mathfrak{c}_+)} \mathrm{M}^{\boldsymbol{G}}(U, \mathfrak{c}; E) \Big/ \left(\{L_x f - f\}_{x \in F^{\times,+}}\right).$$

This recovers classical spaces of Hilbert modular forms.

4.6 Integral models of Hilbert modular varieties

The next step is to try to replicate the constructions of Hilbert moduli varieties and their toroidal compactifications over $\mathbb{Z}$. We start by outlining the construction of integral models of Shimura varieties for the group $\boldsymbol{G}^*$.
The relevant moduli problem, over a base scheme S, is the classification of abelian schemes with real multiplication by $\mathcal{O}_F$, level structure U, and polarisation module $\mathfrak{c}$ as above, *subject to the additional condition* that the natural homomorphism $\mathcal{A} \otimes_{\mathcal{O}_F} \mathrm{Hom}^{\mathrm{sym}}_{S \otimes_{\mathbb{Z}} \mathcal{O}_F}(\mathcal{A}, \mathcal{A}^\vee) \to \mathcal{A}^\vee$, given by $(a, \psi) \mapsto \psi(a)$, is an isomorphism. This is the

Deligne–Pappas condition.

We start with a refresher on integral models of modular curves. Recall from section 1 the geometry of the integral models $\mathcal{X}_0(p)$ from [Deligne–Rapoport, §V]. The moduli stack $\mathcal{X}_0(p)$ classifies generalised elliptic curves E equipped with a cyclic subgroup C of order p which meets all irreducible components of each geometric fibre of E [Deligne–Rapoport, Théorème V.1.6]. There are three options for C.

- The multiplicative case: C is étale locally isomorphic to μ_p.
- The étale case: C is is étale locally isomorphic to $\mathbb{Z}/p\mathbb{Z}$.
- The supersingular case: C is étale locally isomorphic to α_p.

Note that the third situation only makes sense over the locus in which $p = 0$, whereas the first two conditions are equivalent where p is invertible. The first two conditions are Cartier dual because of the Weil pairing: given (E, C) with C of étale type, $(E/C, E[p]/C)$ is of multiplicative type, and vice-versa; the third case is self-dual.
The geometry of $\mathcal{X}_0(p)_{\mathbb{Z}_p}$ is then dictated by these types, as demonstrated by the illustration of $\mathcal{X}_0(p)_{\mathbb{F}_p}$ provided in section 1.
The structural morphism that forgets C behaves differently depending on whether C is étale or multiplicative. On $\mathcal{X}_0(p)^{\mathrm{mult}}_{\mathbb{F}_p}$ it restricts to an isomorphism (the connected-étale sequence provides a unique multiplicative C on an ordinary elliptic curve), whereas on $\mathcal{X}_0(p)^{\text{ét}}_{\mathbb{F}_p}$ it is inseparable of degree p (by duality, it corresponds to the morphism $\mathcal{X}_0(p)^{\mathrm{mult}}_{\mathbb{F}_p} \to \mathcal{X}(1)^{\mathrm{ord}}$ given by quotienting by C). In fact, as noted in section 1, it is the relative Frobenius morphism of $\mathcal{X}_0(1)_{\mathbb{F}_p}$.

Inspired by this situation, when dealing with Hilbert moduli varieties we will restrict to multiplicative level structures. Classifying Hilbert–Blumenthal abelian varieties, the correct definition of a multiplicative level $\mathfrak{n}$ structure on $\mathcal{A}$, for an ideal $\mathfrak{n} \subseteq \mathcal{O}_F$, consists of a homomorphism of group schemes $\psi\colon \mu_{\mathfrak{n}} \to \mathcal{A}[\mathfrak{n}]$, where $\mu_{\mathfrak{n}} := \left(\mathbb{G}_{\mathrm{m}} \otimes_{\mathbb{Z}} \mathfrak{d}_F^{-1}\right)[\mathfrak{n}]$ is the Cartier dual to the constant group scheme $\mathfrak{n}^{-1}\mathcal{O}_F/\mathcal{O}_F$.[76] We will call this a $\mu_{\mathfrak{n}}$-level structure.
The moduli problem for polarised Hilbert–Blumenthal abelian varieties (satisfying the Deligne–Pappas condition) equipped with a $\mu_{\mathfrak{n}}$-level structure is representable by an algebraic stack over $\mathbb{Z}$, which is a scheme for $\mathbf{N}^F_{\mathbb{Q}}\mathfrak{n} \geqslant 4$ [Kisin–Lai, §1.3]. We will denote this stack by $\mathcal{M}(\mu_{\mathfrak{n}}, \mathfrak{c})$.
One possible justification for the use of $\mu_{\mathfrak{n}}$-level structures is to be found by recalling the classical situation we just discussed; forgetting a multiplicative level structure seems like a more natural process than forgetting an étale level structure, as in seen by the difference in the degree of the forgetful maps. In any case, we can recover the other situation by Cartier duality.
There is, however, one additional subtlety in this context: if p ramifies in F, then the special fibre $\mathcal{M}(U(1), \mathfrak{c})_{\mathbb{F}_p}$ is singular in codimension 2 [Deligne–Pappas, Théorème 2.2]. To avoid this additional complication, we can restrict to the smooth locus of $\mathcal{M}(\mu_{\mathfrak{n}}, \mathfrak{c})$; this is usually called the **Rapoport locus**, which we will denote $\mathcal{M}_{\mathrm{R}}(\mu_{\mathfrak{n}}, \mathfrak{c})$.

4.7 Arithmetic toroidal compactifications of Hilbert moduli schemes

We now turn to the question of how compactifications of $\mathcal{M}(\mu_{\mathfrak{n}}, \mathfrak{c})$ relate to their defining moduli problems. The non-properness of $\mathcal{M}(\mu_{\mathfrak{n}}, \mathfrak{c})$ is explained by the existence of discrete valuation rings R, together with a Hilbert–Blumenthal abelian variety A over $K = \mathrm{Frac}(R)$ that does not extend to a Hilbert–Blumenthal abelian scheme over R. To compactify $\mathcal{M}(\mu_{\mathfrak{n}}, \mathfrak{c})$ we thus need to involve degenerating abelian varieties. The miracle is that it suffices to consider semistable degenerations, by Grothendieck's monodromy theorem for abelian varieties [SGA 7_{I}, Exposé IX, Théorème 3.6], which guarantees potential semistability of abelian varieties.

[76] Warning: for an integer n, we have $\mu_{(n)} = \mu_n \otimes_{\mathbb{Z}} \mathfrak{d}_F^{-1}$.

4.7.1 1-motives

Let us now introduce 1-**motives**, which concisely capture the structure of these semistable abelian schemes. This notion is based on the Hodge-theoretic idea that, whereas abelian varieties correspond to Hodge type $\{(0,-1),(-1,0)\}$, they can degenerate into objects with Hodge type $\{(0,0),(0,-1),(-1,0),(-1,-1)\}$. To account for this, Deligne defines 1-motives as follows [Deligne. Hodge III, Définition 10.1.2, Variante 10.1.10]

Definition 4.4 — A 1-motive over S is given by the data of:

- a semiabelian variety G over S with constant toric rank,
- a locally constant étale sheaf of finite free $\mathbb{Z}$-modules Y on S, and
- a morphism $u\colon Y \to G$ of étale sheaves of abelian groups on S.

The 1-motive attached to this data is the object $M = [Y \xrightarrow{u} G]$ in the (bounded) derived category of complexes of étale sheaves of abelian groups on S. ◇

Here, we are meant to understand G as extension of an abelian variety A/S by a torus T/S; then A accounts for Hodge types $\{(0,-1),(-1,0)\}$, T for $\{(-1,-1)\}$ and Y for $\{(0,0)\}$.
To justify the analogy more fully, note that when working over $\mathbb{C}$, just as an abelian variety is uniquely determined by its (torsion-free) polarisable $\mathbb{Z}$-Hodge structure of type $\{(-1,0),(0,-1)\}$, a 1-motive is uniquely determined by its (torsion-free) mixed $\mathbb{Z}$-Hodge structure of type $\{(0,0),(0,-1),(-1,0),(-1,-1)\}$ such that the (-1)-th graded piece of this mixed Hodge structure with respect to the weight filtration is polarisable (this corresponds precisely to the polarisability of the abelian variety occurring as the abelian part of G in the definition of a 1-motive) [Deligne. Hodge III, Construction 10.1.3].

The next step is to understand how these 1-motives lead to degenerating abelian varieties. We start with the following definition [Stroh. Compactifications, Définition 1.3.1] :

Definition 4.5 — Let S be a normal scheme, and $U \hookrightarrow S$ the inclusion of an open dense subscheme. A Mumford 1-motive over $U \hookrightarrow S$ consists of:

- a semiabelian variety G over S with constant toric rank,
- a locally constant étale sheaf of finite free $\mathbb{Z}$-modules Y on S, and
- a 1-motive $[Y_U \xrightarrow{u} G_U]$ over U. ◇

We can then use Mumford 1-motives to describe semiabelian degenerations; this is done by using Mumford, Raynaud and Tate's theory of uniformisation of abelian varieties with semistable reduction.
The basic result is as follows [Faltings–Chai, Corollary III.7.2][77] :

Theorem 4.6 — Let R be a Noetherian excellent domain, complete and separated with respect to a radical ideal I. Let $S = \operatorname{Spec}(R)$, $V = \operatorname{Spec}(R/I)$, $U = S \smallsetminus V$.
Mumford's construction yields an equivalence of categories between:

- the category of polarised Mumford 1-motives over $U \hookrightarrow S$, such that the polarisation is positive with respect to V,[78]
- the category of semiabelian schemes G/S, such that G_U is an abelian scheme, equipped with a given polarisation, and such that the toric part of G_V is isotrivial. ◇□

[77] With small precisions given by [Stroh. Compactifications, Remarque 1.3.3.2, Théorème 1.3.3.3].
[78] See [Stroh. Compactifications, Définition 1.3.3.1] for the precise meaning of this positivity condition.

We can be more specific in the case that interests us, that of Hilbert–Blumenthal abelian schemes. We add the data of an $\mathcal{O}_F$-action; this simplifies the behaviour considerably, as we then only need to consider degenerations with full toric rank. The previous theorem then reduces to the following result, as explained in [Rapoport. Compactifications, §2.8, §2.9] [Andreatta–Goren, §6.1] :

Theorem 4.7 — Let S, U and V be as above. Choose a fractional ideal $\mathfrak{c} \subset F$ with notion of positivity $\mathfrak{c}_+$ (in view of considering $(\mathfrak{c}, \mathfrak{c}_+)$-polarised objects). Mumford's construction yields an equivalence of categories between:

- the category of degenerating Mumford 1-motives over $U \hookrightarrow S$, given by the data of:
  - a (purely toric) semiabelian variety with $\mathcal{O}_F$-action over S, $G = \mathbb{G}_{m,S} \otimes_{\mathbb{Z}} \mathfrak{a}^\vee$, for $\mathfrak{a}$ a fractional ideal of F,
  - a constant sheaf of abelian groups $\mathfrak{b}$, for $\mathfrak{b}$ a fractional ideal of F, and
  - a 1-motive over U, given by a morphism

    $$q\colon \mathfrak{b} \longrightarrow \mathbb{G}_{m,U} \otimes_{\mathbb{Z}} \mathfrak{a}^\vee$$

    of étale sheaves of $\mathcal{O}_L$-modules over U,

  such that:

  - $(\mathfrak{c}, \mathfrak{c}_+) \cong \big(\mathfrak{a}\mathfrak{b}^{-1}, (\mathfrak{a}\mathfrak{b}^{-1})_+\big)$,
  - the following degeneration condition is satisfied: given $m = ab \in (\mathfrak{a}\mathfrak{b})_+$, the element $\chi_a(q(b))$ (where χ_a is the character of $\mathbb{G}_m \otimes_{\mathbb{Z}} \mathfrak{a}^\vee$ associated with $a \in \mathfrak{a}$), considered as an element of the fraction field $K = \mathrm{Frac}(R)$, in fact lies in the ideal I,

  and,

- the category of semiabelian schemes G over S equipped with an $\mathcal{O}_F$-action, such that G_U is an abelian scheme with a given $\mathfrak{c}$-polarisation, and $G_V \cong \mathbb{G}_{m,V} \otimes_{\mathbb{Z}} \mathfrak{a}^\vee$. ◇□

We explain briefly why we should expect the polarisation module in the above theorem to be $\mathfrak{c} = \mathfrak{a}\mathfrak{b}^{-1}$: $\mathfrak{a}\mathfrak{b}^{-1} = \mathrm{Hom}_{\mathcal{O}_F}(\mathfrak{b}, \mathfrak{a}) = \mathrm{Hom}_{\mathcal{O}_F}(\mathfrak{a}^\vee, \mathfrak{b}^\vee)$, so that elements of $\mathfrak{a}\mathfrak{b}^{-1}$ correspond to morphisms

$$\mathbb{G}_m \otimes_{\mathbb{Z}} \mathfrak{a}^\vee \to \mathbb{G}_m \otimes_{\mathbb{Z}} \mathfrak{b}^\vee,$$

and then to morphisms of 1-motives with $\mathcal{O}_F$-action from

$$q\colon \mathfrak{b} \longrightarrow \mathbb{G}_{m,U} \otimes_{\mathbb{Z}} \mathfrak{a}^\vee$$

to the dual 1-motive

$$q\colon \mathfrak{a} \longrightarrow \mathbb{G}_{m,U} \otimes_{\mathbb{Z}} \mathfrak{b}^\vee.$$

We will write the semiabelian scheme G appearing in the above theorem, associated with its corresponding $\mathfrak{c}$-polarised degenerating Mumford 1-motive as in the theorem, as

$$(\mathbb{G}_{m,S} \otimes_{\mathbb{Z}} \mathfrak{a}^\vee)/q^{\mathfrak{b}},$$

where by abuse of notation we write $q^{\mathfrak{b}}$ for $q(\mathfrak{b})$.

4.7.2 Tate semiabelian schemes

To construct arithmetic toroidal compactifications, Rapoport's idea is to explictly construct what one expects to be the formal neighbourhoods of cusps using the previously described Hilbert–Blumenthal semiabelian schemes, and glue these formal neighbourhoods onto $\mathcal{M}(\mu_{\mathfrak{n}}, \mathfrak{c})$, as described in [Rapoport. Compactifications, Introduction].

Starting with fractional ideals of F, $\mathfrak{a}$ and $\mathfrak{b}$, with $\left(\mathfrak{a}\mathfrak{b}^{-1}, (\mathfrak{a}\mathfrak{b}^{-1})_+\right) \cong (\mathfrak{c}, \mathfrak{c}_+)$, we choose a complete fan Σ of salient polyhedral rational convex cones of the dual cone of $((\mathfrak{a}\mathfrak{b})_{\mathbb{R}})_+$ which is invariant under the action of T_U, and with only finitely many cones modulo this action.
Writing S for the torus over $\mathbb{Z}$ with character group $\mathfrak{a}\mathfrak{b}$, attached to Σ is the toric embedding

$$S \hookrightarrow S_\Sigma,$$

obtained by gluing the various S_σ as in section 4.3.2.
If we complete each affine scheme S_σ along $S_\sigma \setminus S$, yielding $S_\sigma^\wedge$, we can apply the results of section 4.7.1 to obtain corresponding $\mathfrak{c}$-polarised Hilbert–Blumenthal semiabelian schemes over $S_\sigma^\wedge$:

$$\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q)_\sigma := \left(\mathbb{G}_{\mathrm{m},S_\sigma^\wedge} \otimes_{\mathbb{Z}} \mathfrak{a}^\vee\right)/q^{\mathfrak{b}}.$$

These are the Tate Hilbert–Blumenthal semiabelian varieties over $S_\sigma^\wedge = \mathrm{Spf}(\mathbb{Z}[\![\sigma^\vee \cap \mathfrak{a}\mathfrak{b}]\!])$.

To make use of these in studying compactifications of $\mathcal{M}(\mu_{\mathfrak{n}}, \mathfrak{c})$, we also need to add level structures. Starting with an ideal $\mathfrak{n} \subseteq \mathcal{O}_F$, tensoring the exact sequence

$$0 \longrightarrow \mathfrak{a}^{-1} \longrightarrow \mathfrak{n}^{-1}\mathfrak{a}^{-1} \longrightarrow \mathfrak{n}^{-1}\mathfrak{a}^{-1}/\mathfrak{a}^{-1} \longrightarrow 0$$

with $\mathbb{G}_{\mathrm{m}} \otimes_{\mathbb{Z}} \mathfrak{d}_F^{-1}$, we obtain the exact sequence

$$0 \longrightarrow \mathrm{Tor}_1^{\mathcal{O}_F}\left(\mathbb{G}_{\mathrm{m}} \otimes_{\mathbb{Z}} \mathfrak{d}_F^{-1}, \mathfrak{n}^{-1}\mathfrak{a}^{-1}/\mathfrak{a}^{-1}\right) \longrightarrow \mathbb{G}_{\mathrm{m}} \otimes_{\mathbb{Z}} \mathfrak{a}^\vee \longrightarrow \mathbb{G}_{\mathrm{m}} \otimes_{\mathbb{Z}} \mathfrak{n}^{-1}\mathfrak{a}^\vee \longrightarrow 0.$$

If $\mathfrak{n}$ is coprime to $\mathfrak{a}$, the $\mathcal{O}_F$-modules $\mathfrak{n}^{-1}\mathfrak{a}^{-1}/\mathfrak{a}^{-1}$ and $\mathfrak{n}^{-1}\mathcal{O}_F/\mathcal{O}_F$ are isomorphic, so that $\mathrm{Tor}_1^{\mathcal{O}_F}\left(\mathbb{G}_{\mathrm{m}} \otimes_{\mathbb{Z}} \mathfrak{d}_F^{-1}, \mathfrak{n}^{-1}\mathfrak{a}^{-1}/\mathfrak{a}^{-1}\right) \cong \mu_{\mathfrak{n}}$. Upon quotienting by $\mathfrak{b}$, we thus arrive at the short exact sequence

$$0 \longrightarrow \mu_{\mathfrak{n}} \longrightarrow \mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q)_\sigma[\mathfrak{n}] \longrightarrow \mathfrak{n}^{-1}\mathfrak{b}/\mathfrak{b} \longrightarrow 0,$$

which is naturally split.
This shows us that the data of an isomorphism $\lambda_1 \colon \mathfrak{n}^{-1}\mathcal{O}_F/\mathcal{O}_F \xrightarrow{\sim} \mathfrak{n}^{-1}\mathfrak{a}^{-1}/\mathfrak{a}^{-1}$ provides $\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q)_\sigma$ with a $\mu_{\mathfrak{n}}$-level structure, which we call an **unramified** level structure.
Another option is possible: if we allow ourselves sufficiently many roots of unity, we can instead directly choose an isomorphism $\lambda_2 \colon \mu_{\mathfrak{n}} \xrightarrow{\sim} \mathfrak{n}^{-1}\mathfrak{b}/\mathfrak{b}$, which also yields a $\mu_{\mathfrak{n}}$-level structure on $\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q)_\sigma$, which we say is **ramified**.

We also want to understand sheaves of differentials on $\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q)$, as these will be needed to extend the automorphic vector bundles to the compactifications of $M(U, \mathfrak{c})$. We can straightforwardly use the invariant differential $\frac{\mathrm{d}t}{t}$ of $\mathbb{G}_{\mathrm{m}}$ to see that we have:

$$\omega_{\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q)_\sigma/S_\sigma^\wedge} \cong \left(\mathcal{O}_{S_\sigma^\wedge} \otimes_{\mathbb{Z}} \mathfrak{a}\right)\frac{\mathrm{d}t}{t}.$$

4.7.3 Arithmetic toroidal compactifications

With the Tate semiabelian schemes in hand, we can now describe the arithmetic toroidal compactifications of $\mathcal{M}(\mu_{\mathfrak{n}}, \mathfrak{c})$, together with the corresponding universal Hilbert–Blumenthal semiabelian scheme.
To begin, we recall the analytic description of the universal Hilbert–Blumenthal abelian variety over $M(U, \mathfrak{c})$, following [Dimitrov–Tilouine, §2]. Given a congruence subgroup $\Gamma \subset \mathrm{SL}^*(\mathcal{O}_F \oplus \mathfrak{c}^\vee)$, we can think of the connected Shimura varieties $\Gamma\backslash\mathcal{H}_F$ as parametrising lattices (with additional level structure). We then want to perform the complex-analytic quotient of $F \otimes_{\mathbb{Q}} \mathbb{C}$ by the varying lattice parametrised by $\Gamma\backslash\mathcal{H}_F$, giving a bundle of abelian varieties over $\Gamma\backslash\mathcal{H}_F$. This is achieved by taking the following double quotient:

$$\Gamma\backslash\mathcal{H}_F \times (F \otimes_{\mathbb{Q}} \mathbb{C})/\left(\mathcal{O}_F \oplus \mathfrak{c}^\vee\right),$$

with the left action of Γ on $\mathcal{H}_F \times (F \otimes_{\mathbb{Q}} \mathbb{C})$ given by

$$\gamma \cdot (z, v) = (\gamma \cdot z, j(\gamma, z)v),$$

where $j(\gamma, z)$ is the usual automorphy factor $j(\gamma, z) = (c_\tau z + d_\tau)_\tau$ for $\gamma = \left(\begin{pmatrix} a_\tau & b_\tau \\ c_\tau & d_\tau \end{pmatrix}\right)_\tau$, whereas the right action of $\mathcal{O}_F \oplus \mathfrak{c}^\vee$ is given by

$$(z, v) \cdot (m, n) = (z, v + mz + n).$$

This ensures that the fiber above a given point $[z]$ in $\Gamma\backslash\mathcal{H}_F$ is the abelian variety $\mathcal{A}_z = (F \otimes_{\mathbb{Q}} \mathbb{C})/(z\mathcal{O}_F \oplus \mathfrak{c}^\vee)$. We now want to concentrate our attention to the behaviour of $\mathcal{A}$ around cusps. As in the case of modular curves, by the results of sections 4.7.1 and 4.7.2, we expect $\mathcal{A}$ to look like, around the cusp labelled by $(\mathfrak{a}, \mathfrak{b})$,

$$\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q) := (\mathbb{G}_{\mathrm{m}} \otimes_{\mathbb{Z}} \mathfrak{a}^\vee)\, / q^{\mathfrak{b}},$$

the Tate Hilbert–Blumenthal semiabelian variety over $\mathbb{Z}[(\mathfrak{a}\mathfrak{b})_+]$.
Around the cusp ∞, the universal Hilbert–Blumenthal abelian variety $\mathcal{A}$ then fits into the following commutative diagram:

$$\begin{tikzcd}
\mathcal{A} \arrow[d, "\pi"'] & B_U\backslash W(\infty,r)\times\big(F\otimes_{\mathbb{Q}}\mathbb{C}\big)/(\mathcal{O}_F\oplus\mathfrak{c}^\vee) \arrow[l, hook'] \arrow[d] & V_U\backslash W(\infty,r)\times\big(F\otimes_{\mathbb{Q}}\mathbb{C}\big)/(\mathcal{O}_F\oplus\mathfrak{c}^\vee) \arrow[l] \arrow[r, hook] \arrow[d] & \big(\mathbb{G}_{\mathrm{m}}\otimes_{\mathbb{Z}}\mathfrak{c}^\vee\times\mathbb{G}_{\mathrm{m}}\otimes_{\mathbb{Z}}\mathfrak{c}^\vee\big)\big/(q_z)^{\mathcal{O}_F} \arrow[d] \\
M(U,\mathfrak{c}) & B_U\backslash W(\infty,r) \arrow[l, hook'] & V_U\backslash W(\infty,r) \arrow[l, "T_U\backslash-"'] \arrow[r, hook, "q"] & \mathbb{G}_{\mathrm{m}}\otimes_{\mathbb{Z}}\mathfrak{c}^\vee
\end{tikzcd}$$

Here we are writing the two coordinates on $\mathbb{G}_{\mathrm{m}} \otimes_{\mathbb{Z}} \mathfrak{c}^\vee \times \mathbb{G}_{\mathrm{m}} \otimes_{\mathbb{Z}} \mathfrak{c}^\vee$ as (q_z, q_v), and the quotient $\big(\mathbb{G}_{\mathrm{m}} \otimes_{\mathbb{Z}} \mathfrak{c}^\vee \times \mathbb{G}_{\mathrm{m}} \otimes_{\mathbb{Z}} \mathfrak{c}^\vee\big)\big/(q_z)^{\mathcal{O}_F}$ is to be understood with respect to the action

$$(q_z, q_v) \cdot b = \left(q_z, q_v \cdot (q_z)^b\right)$$

for $b \in \mathcal{O}_F$. On the other hand, the action of T_U also extends to $\mathbb{G}_{\mathrm{m}} \otimes_{\mathbb{Z}} \mathfrak{c}^\vee \times \mathbb{G}_{\mathrm{m}} \otimes_{\mathbb{Z}} \mathfrak{c}^\vee$, being given by

$$\begin{pmatrix} \varepsilon & 0 \\ 0 & \varepsilon^{-1} \end{pmatrix} \cdot (q_z, q_v) = \left(q_z^{\varepsilon^2}, q_v^{\varepsilon}\right).$$

Over a given point q_z of $\mathbb{G}_{\mathrm{m}} \otimes_{\mathbb{Z}} \mathfrak{c}^\vee$, one recovers the analytic quotient $\mathbb{G}_{\mathrm{m}} \otimes_{\mathbb{Z}} \mathfrak{c}^\vee/(q_z)^{\mathcal{O}_F}$. A similar description can be given around other cusps [Dimitrov–Tilouine, Définition 2.11].

The main result in the theory of arithmetic toroidal compactifications of Hilbert modular schemes is then as follows [Chai. Compactifications, Theorem 3.6] :

Theorem 4.8 — Let Σ be a collection of fans adapted[79] to $\mathcal{M}(\mu_{\mathfrak{n}}, \mathfrak{c})$. There is an embedding (of algebraic stacks)

$$j\colon \mathcal{M}(\mu_{\mathfrak{n}}, \mathfrak{c}) \hookrightarrow \overline{\mathcal{M}}_\Sigma(\mu_{\mathfrak{n}}, \mathfrak{c}),$$

and a Hilbert–Blumenthal semiabelian scheme $\widetilde{\pi}\colon \overline{\mathcal{A}} \to \overline{\mathcal{M}}_\Sigma(\mu_{\mathfrak{n}}, \mathfrak{c})$, extending $\pi\colon \mathcal{A} \to \mathcal{M}(\mu_{\mathfrak{n}}, \mathfrak{c})$, such that the formal completion of $\mathcal{M}(\mu_{\mathfrak{n}}, \mathfrak{c})$ along its boundary matches up with the construction of section 4.3.3. Moreover, over this formal completion, $\widetilde{\mathcal{A}}$ is described by Mumford's construction (see sections 4.7.1 and 4.7.2). $\diamond\square$

4.8 Integral Hilbert modular forms

Now that we have integral models of the Shimura varieties, we also need to extend the definition of the automorphic vector bundles. We again can use the universal Hilbert–Blumenthal abelian scheme $\pi\colon \mathcal{A} \to \mathcal{M}(\mu_{\mathfrak{n}}, \mathfrak{c})$ to define, as previously, the sheaves $\omega(\kappa)$, using the sheaf $\omega_{\mathcal{A}}$. Recall however that we needed to know that ω was a locally free $\mathcal{O}_{\mathcal{M}(\mu_{\mathfrak{n}},\mathfrak{c})} \otimes_{\mathbb{Z}} \mathcal{O}_F$-module of rank 1. This is no longer necessarily true: while ω is a locally free sheaf of $\mathcal{O}_{\mathcal{M}(\mu_{\mathfrak{n}},\mathfrak{c})}$-modules equipped with an action of $\mathcal{O}_F$, it is not necessarily locally free as a sheaf of

[79] A collection of fans is adapted if each fan satisfies the conditions with respect to the action of the units at its corresponding cusp, and if the fans at different cusps are chosen compatibly with the action of $\boldsymbol{G}^*(\mathbb{Q})$. These exist [Chai. Compactifications, §3.3].

$\mathcal{O}_{\mathcal{M}(\mu_{\mathfrak{n}},\mathfrak{c})} \otimes_{\mathbb{Z}} \mathcal{O}_F$-modules. The Rapoport locus $\mathcal{M}_{\mathrm{R}}(\mu_{\mathfrak{n}},\mathfrak{c})$ of $\mathcal{M}(\mu_{\mathfrak{n}},\mathfrak{c})$ is in fact the largest open substack of $\mathcal{M}(\mu_{\mathfrak{n}},\mathfrak{c})$ over which ω *is* locally free as an $\mathcal{O}_{\mathcal{M}(\mu_{\mathfrak{n}},\mathfrak{c})} \otimes_{\mathbb{Z}} \mathcal{O}_F$-module. To define integral models of the automorphic vector bundles $\omega(\kappa)$, we are thus forced to use the Rapoport locus.

We've seen in sections 4.7.2 and 4.7.3 how to extend the universal Hilbert–Blumenthal abelian scheme $\pi\colon \mathcal{A} \to \mathcal{M}(\mu_{\mathfrak{n}},\mathfrak{c})$ to a semiabelian scheme $\widetilde{\pi}\colon \overline{\mathcal{A}}_{\Sigma} \to \overline{\mathcal{M}}_{\Sigma}(\mu_{\mathfrak{n}},\mathfrak{c})$, so we can then use the sheaf of canonical differentials $\omega_{\overline{\mathcal{A}}}$ and extend ω to the whole of $\overline{\mathcal{M}}_{\Sigma}(\mu_{\mathfrak{n}},\mathfrak{c})$ by defining $\omega = e^*\omega_{\overline{\mathcal{A}}}$, where e is the identity section of $\widetilde{\pi}$.
We finally can define integral spaces of Hilbert modular forms for $\boldsymbol{G}^*$:

$$\mathrm{M}_{\kappa}^{\boldsymbol{G}^*}(\mu_{\mathfrak{n}},\mathfrak{c};\mathcal{O}_E) := \mathrm{H}^0\left(\overline{\mathcal{M}}_{\Sigma,\mathrm{R}}(\mu_{\mathfrak{n}},\mathfrak{c})_{\mathcal{O}_E}, \omega(\kappa)\right).$$

We can also proceed similarly for $\boldsymbol{G}$, defining

$$\mathrm{M}_{(\nu,w)}^{\boldsymbol{G}}(\mu_{\mathfrak{n}},\mathfrak{c};\mathcal{O}_E) := \mathrm{H}^0\left(\overline{\mathcal{M}}_{\Sigma,\mathrm{R}}^{\boldsymbol{G}}(\mu_{\mathfrak{n}},\mathfrak{c})_{\mathcal{O}_E}, \omega(\nu,w)\right),$$

and the corresponding spaces $\mathrm{M}_{(\nu,w)}^{\boldsymbol{G}}(\mu_{\mathfrak{n}};\mathcal{O}_E)$ defined as in section 4.5.
Crucially, these spaces do not depend on the choice of Σ, because of the following result [Chai. Compactifications, Theorem 4.3.(i)] :

Theorem 4.9 (Koecher principle) — If $[F:\mathbb{Q}] > 1$, every global section of ω on $\mathcal{M}(\mu_{\mathfrak{n}},\mathfrak{c})$ naturally extends to the toroidal compactifications $\overline{\mathcal{M}}_{\Sigma}(\mu_{\mathfrak{n}},\mathfrak{c})$. ◇□

Now, from Theorem 4.8 we know what the formal neighbourhoods of cusps look like. As the universal Hilbert–Blumenthal semiabelian variety is given around (formal neighbourhoods of) the cusps by the Tate semiabelian varieties $\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q)_{\Sigma}$ (once they are given appropriate level structures), this means we are able to (in the Katz formalism) evaluate Hilbert modular forms at these Tate semiabelian varieties, obtaining power series in the relevant completed local rings.
This gives q-expansions of Hilbert modular forms at cusps. We then have the following fundamental principle [Chai. Compactifications, Theorem 4.3.(ii)] :

Theorem 4.10 (q-expansion principle) — The q-expansion homomorphism at a cusp C has kernel consisting precisely of those sections which vanish on the connected component of $\overline{\mathcal{M}}_{\Sigma}(\mu_{\mathfrak{n}},\mathfrak{c})$ containing C. Moreover, a section f of $\omega(\kappa)$ on $M(U_0(\mathfrak{n}),\mathfrak{c})_E$ is defined over $\mathcal{M}_{\mathrm{R}}(\mu_{\mathfrak{n}},\mathfrak{c})_{\mathcal{O}_E}$ if and only the q-expansions of f at all cusps C of $M(U_0(\mathfrak{n}),\mathfrak{c})_E$ have integral coefficients (equivalently, if each connected component of $M(U_0(\mathfrak{n}),\mathfrak{c})_E$ contains a cusp C at which the q-expansion of f is integral). ◇□

5 Hecke operators on Hilbert modular forms

We want to prove an integrality result for the action of Hecke operators on Hilbert modular forms. To do this, we would like to define Hecke correspondences on integral models of Hilbert–Blumenthal abelian varieties.
To have at hand the automorphic vector bundles $\omega(\kappa)$, we need to work with the Rapoport loci $\mathcal{M}_{\mathrm{R}}(\mu_{\mathfrak{n}},\mathfrak{c})$ and their toroidal compactifications $\overline{\mathcal{M}}_{\Sigma,\mathrm{R}}(\mu_{\mathfrak{n}},\mathfrak{c})$. We begin by geometrically defining integral avatars of Hecke operators by considering a moduli problem where an extra level structure is introduced. Unfortunately, this can only be done at primes that do not divide the level $\mathfrak{n}$; we will tackle the primes dividing $\mathfrak{n}$ separately in section 5.3.

5.1 Integral Hecke operators for $\boldsymbol{G}^*$

We start by defining the Hecke correspondences. Choose a prime ideal $\mathfrak{p} \subset \mathcal{O}_F$ that does not divide $\mathfrak{n}$,[80] and consider the moduli problem of triples

$$\Big((\mathcal{A}_1,\iota_1,\psi_1,\lambda_1),(\mathcal{A}_2,\iota_2,\psi_2,\lambda_2),\alpha\Big) \in \mathcal{M}(\mu_{\mathfrak{n}},\mathfrak{c}) \times \mathcal{M}(\mu_{\mathfrak{n}},\mathfrak{p}\mathfrak{c}) \times \mathrm{Hom}(\mathcal{A}_1,\mathcal{A}_2),$$

[80] This condition is necessary; otherwise we would need to introduce some kind of transversality condition between the two level structures, which can be difficult to formulate correctly when working over $\mathbb{Z}$ instead of over $\mathbb{Q}$.

where $\alpha\colon \mathcal{A}_1 \to \mathcal{A}_2$ is an isogeny of degree $\mathbf{N}_{\mathbb{Q}}^F \mathfrak{p}$, required to be compatible with the $\mathcal{O}_F$ action (via ι_1 and ι_2), the polarisations (ψ_1 and ψ_2) and the level structures (λ_1 and λ_2), with $\ker(\alpha) \subset \mathcal{A}_1[\mathfrak{p}]$ killed by $\mathfrak{p}$. This moduli problem is representable by an algebraic stack over $\mathbb{Z}$, which we will write $\mathcal{Y}(\mu_{\mathfrak{n}}, \mathfrak{c}, \mathfrak{a}\mathfrak{c})$. It comes equipped with two projection maps, $p_1\colon \mathcal{Y}(\mu_{\mathfrak{n}}, \mathfrak{c}, \mathfrak{a}\mathfrak{c}) \to \mathcal{M}(\mu_{\mathfrak{n}}, \mathfrak{c})$ and $p_2\colon \mathcal{Y}(\mu_{\mathfrak{n}}, \mathfrak{c}, \mathfrak{a}\mathfrak{c}) \to \mathcal{M}(\mu_{\mathfrak{n}}, \mathfrak{p}\mathfrak{c})$. Unfortunately these two morphisms are not necessarily finite; we restrict to the ordinary loci $\mathcal{M}(\mu_{\mathfrak{n}}, \mathfrak{c})^{\mathrm{ord}} \subseteq \mathcal{M}_{\mathrm{R}}(\mu_{\mathfrak{n}}, \mathfrak{c})$. Over these loci, the morphisms p_1 and p_2 are finite and flat [Emerton–Reduzzi–Xiao, Proposition 3.7], and a naive Hecke operator $\mathrm{T}_{\mathfrak{p}}^{\mathrm{naive}}$ can then be defined as the composite arrow in the following diagram:

$$\begin{tikzcd}
\mathrm{H}^0\left(\mathcal{M}(\mu_{\mathfrak{n}}, \mathfrak{p}\mathfrak{c})^{\mathrm{ord}}_{\mathcal{O}_E}, \omega(\kappa)\right) \arrow[r, "(p_2)^*"] & \mathrm{H}^0\left(\mathcal{Y}(\mu_{\mathfrak{n}}, \mathfrak{c}, \mathfrak{a}\mathfrak{c})^{\mathrm{ord}}_{\mathcal{O}_E}, (p_2)^*\omega(\kappa)\right) \arrow[d, "\alpha^*"] & \\
& \mathrm{H}^0\left(\mathcal{Y}(\mu_{\mathfrak{n}}, \mathfrak{c}, \mathfrak{a}\mathfrak{c})^{\mathrm{ord}}_{\mathcal{O}_E}, (p_1)^*\omega(\kappa)\right) \arrow[r, "\mathrm{tr}(p_1)"'] & \mathrm{H}^0\left(\mathcal{M}(\mu_{\mathfrak{n}}, \mathfrak{c})^{\mathrm{ord}}_{\mathcal{O}_E}, \omega(\kappa)\right).
\end{tikzcd}$$

The key ingredient holding this together is the use of the universal isogeny $\alpha\colon \mathcal{A}_1 \to \mathcal{A}_2$; this explains why we are pulling back by p_2 and pushing forward by p_1, and not the other way around.
Now, as we have defined these Hecke correspondences on the integral models, they are automatically integral. However, the crucial subtlety lies in the possibility of dividing by $\mathbf{N}_{\mathbb{Q}}^F \mathfrak{p}$ while preserving integrality.

Proposition 5.1 — The operator

$$\mathrm{T}_{\mathfrak{p}}^{\mathrm{naive}}\colon \mathrm{H}^0\left(\mathcal{M}(\mu_{\mathfrak{n}}, \mathfrak{p}\mathfrak{c})^{\mathrm{ord}}_{\mathcal{O}_E}, \omega(\kappa)\right) \to \mathrm{H}^0\left(\mathcal{M}_{\mathrm{R}}(\mu_{\mathfrak{n}}, \mathfrak{c})^{\mathrm{ord}}_{\mathcal{O}_E}, \omega(\kappa)\right)$$

has image contained in $\mathbf{N}_{\mathbb{Q}}^F \mathfrak{p} \cdot \mathrm{H}^0\left(\mathcal{M}_{\mathrm{R}}(\mu_{\mathfrak{n}}, \mathfrak{c})^{\mathrm{ord}}_{\mathcal{O}_E}, \omega(\kappa)\right)$. ◇

Proof: It suffices to consider the situation p-adically, where $p \in \mathbb{Z}$ is the unique prime of $\mathbb{Z}$ lying under $\mathfrak{p}$. Abusing notation, we will write $\mathcal{M}(\mu_{\mathfrak{n}}, \mathfrak{c})$ and $\mathcal{Y}(\mu_{\mathfrak{n}}, \mathfrak{c}, \mathfrak{p}\mathfrak{c})$ for the $\mathbb{Z}_p$-integral models of the relevant Hilbert–Blumenthal moduli spaces, and corresponding Rapoport and ordinary loci.
Similarly to the Deligne–Rapoport picture of section 4.6, $\mathcal{Y}(\mu_{\mathfrak{n}}, \mathfrak{c}, \mathfrak{a}\mathfrak{c})^{\mathrm{ord}}$ has two components $\mathcal{Y}(\mu_{\mathfrak{n}}, \mathfrak{c}, \mathfrak{a}\mathfrak{c})^{\mathrm{mult}}$ and $\mathcal{Y}(\mu_{\mathfrak{n}}, \mathfrak{c}, \mathfrak{a}\mathfrak{c})^{\text{ét}}$, corresponding to whether $\ker(\alpha)$ is multiplicative or étale, respectively. Restricted to $\mathcal{Y}(\mu_{\mathfrak{n}}, \mathfrak{c}, \mathfrak{a}\mathfrak{c})^{\mathrm{mult}}$, p_1 is an isomorphism, whereas on $\mathcal{Y}(\mu_{\mathfrak{n}}, \mathfrak{c}, \mathfrak{a}\mathfrak{c})^{\text{ét}}$ it is purely inseparable of degree $|\mathcal{O}_F/\mathfrak{p}| = \mathbf{N}_{\mathbb{Q}}^F \mathfrak{p}$, to be thought of as a $\mathfrak{p}$-Frobenius morphism. Following the method of [Conrad. Moduli, Theorem 4.5.1], we can study these two situations separately.
If $\ker(\alpha)$ is étale, this description of p_1 then implies that

$$\mathrm{tr}(p_1)\colon \mathrm{H}^0\left(\mathcal{Y}(\mu_{\mathfrak{n}}, \mathfrak{c}, \mathfrak{a}\mathfrak{c})^{\mathrm{ord}}, (p_1)^*\omega\right) \longrightarrow \mathrm{H}^0\left(\mathcal{M}(\mu_{\mathfrak{n}}, \mathfrak{c})^{\mathrm{ord}}, \omega\right)$$

is divisible by $\mathbf{N}_{\mathbb{Q}}^F \mathfrak{p}$ [Emerton–Reduzzi–Xiao, Proposition 3.16, Case 2].
On the other hand, if $\ker(\alpha)$ is multiplicative, α is inseparable of degree $\mathbf{N}_{\mathbb{Q}}^F \mathfrak{p}$, and therefore

$$\alpha^*\colon \mathrm{H}^0\left(\mathcal{Y}(\mu_{\mathfrak{n}}, \mathfrak{c}, \mathfrak{a}\mathfrak{c})^{\mathrm{ord}}, (p_2)^*\omega\right) \longrightarrow \mathrm{H}^0\left(\mathcal{Y}(\mu_{\mathfrak{n}}, \mathfrak{c}, \mathfrak{a}\mathfrak{c})^{\mathrm{ord}}, (p_1)^*\omega\right)$$

is divisible by $\mathbf{N}_{\mathbb{Q}}^F \mathfrak{p}$ [Emerton–Reduzzi–Xiao, Proposition 3.16, Case 1].
This shows the required statement with regards to the sheaf ω on $\mathcal{M}(\mu_{\mathfrak{n}}, \mathfrak{p}\mathfrak{c})^{\mathrm{ord}}$; the same method then yields identical results for the sheaves $\omega(\kappa)$ on $\mathcal{M}(\mu_{\mathfrak{n}}, \mathfrak{p}\mathfrak{c})^{\mathrm{ord}}_{\mathcal{O}_E}$, as $\kappa_\tau \geqslant 1$ for all τ. □

We then write $\mathrm{T}_{\mathfrak{p}} = \left(\mathbf{N}_{\mathbb{Q}}^F \mathfrak{p}\right)^{-1} \mathrm{T}_{\mathfrak{p}}^{\mathrm{naive}}$ for the renormalised Hecke operator, which we now know is integral. Recall however that we have assumed that $\mathfrak{p}$ does not divide $\mathfrak{n}$; we will tackle the other case in section 5.3.
In a similar fashion, we can also define a naive Hecke operator $\mathrm{S}_{\mathfrak{p}}^{\mathrm{naive}}$. To do this, we consider the morphism $\alpha_{\mathfrak{p}}\colon \mathcal{M}(\mu_{\mathfrak{n}}, \mathfrak{c}) \to \mathcal{M}(\mu_{\mathfrak{n}}, \mathfrak{p}^2\mathfrak{c})$ induced by

$$(\mathcal{A}, \iota, \psi, \lambda) \mapsto (\mathcal{A} \otimes_{\mathcal{O}_F} \mathfrak{p}, \iota', \mathfrak{p}^2\psi, \lambda'),$$

where ι' and λ' are induced from ι and λ. Pullback along $\alpha_{\mathfrak{p}}$ then induces

$$\mathrm{S}_{\mathfrak{p}}^{\mathrm{naive}} \colon \mathrm{H}^0\left(\mathcal{M}_{\mathrm{R}}(\mu_{\mathfrak{n}}, \mathfrak{p}^2\mathfrak{c})_{\mathcal{O}_E}, \omega(\kappa)\right) \to \mathrm{H}^0\left(\mathcal{M}_{\mathrm{R}}(\mu_{\mathfrak{n}}, \mathfrak{c})_{\mathcal{O}_E}, \omega(\kappa)\right),$$

implicitly using the natural morphism $(\alpha_{\mathfrak{p}})^*\omega(\kappa) \to \omega(\kappa)$. A similar argument to the one made in the proof of Proposition 5.1 shows that the image of $\mathrm{S}_{\mathfrak{p}}^{\mathrm{naive}}$ is divisible by $\mathbf{N}_{\mathbb{Q}}^{F}\mathfrak{p}$. However, the usual normalisation is given by $\mathrm{S}_{\mathfrak{p}} := \left(\mathbf{N}_{\mathbb{Q}}^{F}\mathfrak{p}\right)^{-2} \mathrm{S}_{\mathfrak{p}}^{\mathrm{naive}}$, which is not necessarily integral.

5.2 Integral Hecke operators for G

We now want to do the same thing for $\boldsymbol{G}$. In this case we have moduli stacks associated with $\boldsymbol{G}$, and line bundles $\omega(k,l) = \omega(k) \otimes_{\mathcal{O}_{\mathcal{M}_{\mathrm{R}}^{\boldsymbol{G}}(\mu_{\mathfrak{n}},\mathfrak{c})_{\mathcal{O}_E}}} \delta(l)$. We can again define naive Hecke operators in the same way as above:

$$\begin{tikzcd}
\mathrm{H}^0\left(\mathcal{M}^{\boldsymbol{G}}(\mu_{\mathfrak{n}}, \mathfrak{p}\mathfrak{c})_{\mathcal{O}_E}^{\mathrm{ord}}, \omega(k,l)\right) \arrow[r, "(p_2)^*"] & \mathrm{H}^0\left(\mathcal{Y}^{\boldsymbol{G}}(\mu_{\mathfrak{n}}, \mathfrak{c}, \mathfrak{p}\mathfrak{c})_{\mathcal{O}_E}^{\mathrm{ord}}, (p_2)^*\omega(k,l)\right) \arrow[d, "\alpha^*"] & \\
& \mathrm{H}^0\left(\mathcal{Y}^{\boldsymbol{G}}(\mu_{\mathfrak{n}}, \mathfrak{c}, \mathfrak{p}\mathfrak{c})_{\mathcal{O}_E}^{\mathrm{ord}}, (p_1)^*\omega(k,l)\right) \arrow[r, "\mathrm{tr}(p_1)"'] & \mathrm{H}^0\left(\mathcal{M}^{\boldsymbol{G}}(\mu_{\mathfrak{n}}, \mathfrak{c})_{\mathcal{O}_E}^{\mathrm{ord}}, \omega(k,l)\right).
\end{tikzcd}$$

We now want to establish a similar integrality result as the one given in Proposition 5.1. The subtlety arises from the twists by δ, which complicate the integrality properties of the Hecke operators as noted by Hida [Hida. p-adic Hecke algebras, §4]. We have:

Proposition 5.2 — The operator

$$\mathrm{T}_{\mathfrak{p}}^{\mathrm{naive}} \colon \mathrm{H}^0\left(\mathcal{M}^{\boldsymbol{G}}(\mu_{\mathfrak{n}}, \mathfrak{p}\mathfrak{c})_{\mathcal{O}_E}^{\mathrm{ord}}, \omega(k,l)\right) \longrightarrow \mathrm{H}^0\left(\mathcal{M}^{\boldsymbol{G}}(\mu_{\mathfrak{n}}, \mathfrak{c})_{\mathcal{O}_E}^{\mathrm{ord}}, \omega(k,l)\right)$$

has image contained in $\left(\mathbf{N}_{\mathbb{Q}}^{F}\mathfrak{p}\right) \cdot \prod_{\tau\colon F\hookrightarrow E} \tau(\mathfrak{p})^{l_\tau} \cdot \mathrm{H}^0\left(\mathcal{M}^{\boldsymbol{G}}(\mu_{\mathfrak{n}}, \mathfrak{c})_{\mathcal{O}_E}^{\mathrm{ord}}, \omega(k,l)\right)$. ◇

Proof: The same argument as in the proof of Proposition 5.1 yields the divisibility by $\mathbf{N}_{\mathbb{Q}}^{F}\mathfrak{p}$; running that argument in parallel with the remainder of the following proof will take care of that factor.
Let us now show that the image of

$$\alpha^* \colon \mathrm{H}^0\left(\mathcal{Y}_{\mathrm{R}}^{\boldsymbol{G}}(\mu_{\mathfrak{n}}, \mathfrak{c}, \mathfrak{p}\mathfrak{c})_{\mathcal{O}_E}, (p_2)^*\omega(k,l)\right) \to \mathrm{H}^0\left(\mathcal{Y}_{\mathrm{R}}^{\boldsymbol{G}}(\mu_{\mathfrak{n}}, \mathfrak{c}, \mathfrak{p}\mathfrak{c})_{\mathcal{O}_E}, (p_1)^*\omega(k,l)\right)$$

is contained in $\prod_{\tau\colon F\hookrightarrow E} \tau(\mathfrak{p})^{l_\tau} \cdot \mathrm{H}^0\left(\mathcal{Y}_{\mathrm{R}}^{\boldsymbol{G}}(\mu_{\mathfrak{n}}, \mathfrak{c}, \mathfrak{p}\mathfrak{c})_{\mathcal{O}_E}, \omega(k,l)\right)$.
To show this, we use that the line bundle δ on $\mathcal{M}_{\mathrm{R}}^{\boldsymbol{G}}(\mu_{\mathfrak{n}}, \mathfrak{c})$ is free (not just locally free). Indeed, recall the argument from section 4.5: the Hodge filtration of the universal Hilbert–Blumenthal abelian variety $\mathcal{A}$ is given by the short exact sequence

$$0 \longrightarrow \omega \longrightarrow \mathcal{H}_{\mathrm{dR}}^1\left(\mathcal{A}/\mathcal{M}^{\boldsymbol{G}}(\mu_{\mathfrak{n}}, \mathfrak{c})\right) \longrightarrow \omega^\vee \otimes_{\mathbb{Z}} \mathfrak{c}\mathfrak{d}_F^{-1} \longrightarrow 0.$$

Over the Rapoport locus $\mathcal{M}_{\mathrm{R}}^{\boldsymbol{G}}(\mu_{\mathfrak{n}}, \mathfrak{c})$, ω is an invertible $\mathcal{O}_{\mathcal{M}_{\mathrm{R}}^{\boldsymbol{G}}(\mu_{\mathfrak{n}},\mathfrak{c})} \otimes_{\mathbb{Z}} \mathcal{O}_F$-module, so that

$$\delta = \bigwedge\nolimits^2_{\mathcal{O}_{\mathcal{M}_{\mathrm{R}}^{\boldsymbol{G}}(\mu_{\mathfrak{n}},\mathfrak{c})} \otimes_{\mathbb{Z}} \mathcal{O}_F} \mathcal{H}_{\mathrm{dR}}^1\left(\mathcal{A}/\mathcal{M}_{\mathrm{R}}^{\boldsymbol{G}}(\mu_{\mathfrak{n}}, \mathfrak{c})\right) \cong \mathfrak{c}\mathfrak{d}_F^{-1} \otimes_{\mathbb{Z}} \mathcal{O}_{\mathcal{M}_{\mathrm{R}}^{\boldsymbol{G}}(\mu_{\mathfrak{n}},\mathfrak{c})}.$$

As a consequence,

$$(p_1)^*\delta \cong \mathfrak{c}\mathfrak{d}_F^{-1} \otimes_{\mathbb{Z}} \mathcal{O}_{\mathcal{Y}_{\mathrm{R}}^{\boldsymbol{G}}(\mu_{\mathfrak{n}},\mathfrak{c},\mathfrak{p}\mathfrak{c})}, \qquad (p_2)^*\delta \cong \mathfrak{p}\mathfrak{c}\mathfrak{d}_F^{-1} \otimes_{\mathbb{Z}} \mathcal{O}_{\mathcal{Y}_{\mathrm{R}}^{\boldsymbol{G}}(\mu_{\mathfrak{n}},\mathfrak{c},\mathfrak{p}\mathfrak{c})},$$

and α^* is given by the inclusion of $\mathfrak{p}\mathfrak{c}\mathfrak{d}_F^{-1}$ in $\mathfrak{c}\mathfrak{d}_F^{-1}$ [Emerton–Reduzzi–Xiao, §3.9].
Passing now to $\mathcal{O}_E$, we decompose $\delta = \bigoplus_{\tau\colon F\hookrightarrow E} \delta_\tau$ into a direct sum of line bundles. Taking tensor products to form the line bundles $\delta(l) = \bigotimes_{\tau\colon F\hookrightarrow E} \delta_\tau^{\otimes l_\tau}$, the inclusion $\mathfrak{p}\mathfrak{c}\mathfrak{d}_F^{-1} \subset \mathfrak{c}\mathfrak{d}_F^{-1}$ ensures that the image of

$$\alpha^* \colon \mathrm{H}^0\left(\mathcal{Y}_{\mathrm{R}}^{\boldsymbol{G}}(\mu_{\mathfrak{n}}, \mathfrak{c}, \mathfrak{p}\mathfrak{c})_{\mathcal{O}_E}, (p_2)^*\omega(k,l)\right) \to \mathrm{H}^0\left(\mathcal{Y}_{\mathrm{R}}^{\boldsymbol{G}}(\mu_{\mathfrak{n}}, \mathfrak{c}, \mathfrak{p}\mathfrak{c})_{\mathcal{O}_E}, (p_1)^*\omega(k,l)\right)$$

is indeed contained in $\prod_{\tau:F\hookrightarrow E}\tau(\mathfrak{p})^{l_\tau}\cdot \mathrm{H}^0\left(\mathcal{Y}^{\boldsymbol{G}}_{\mathrm{R}}(\mu_{\mathfrak{n}},\mathfrak{c},\mathfrak{p}\mathfrak{c})_{\mathcal{O}_E},\omega(k,l)\right)$. $\square$

We also get a result for the operators $\mathrm{S}_{\mathfrak{p}}$: the image of

$$\mathrm{S}_{\mathfrak{p}}^{\mathrm{naive}}\colon \mathrm{H}^0\left(\mathcal{M}^{\boldsymbol{G}}(\mu_{\mathfrak{n}},\mathfrak{p}^2\mathfrak{c})^{\mathrm{ord}}_{\mathcal{O}_E},\omega(k,l)\right)\longrightarrow \mathrm{H}^0\left(\mathcal{M}^{\boldsymbol{G}}(\mu_{\mathfrak{n}},\mathfrak{c})^{\mathrm{ord}}_{\mathcal{O}_E},\omega(k,l)\right)$$

lies in $\left(\mathbf{N}^F_{\mathbb{Q}}\mathfrak{p}\right)\cdot\prod_\tau\tau(\mathfrak{p})^{2l_\tau}\cdot\mathrm{H}^0\left(\mathcal{M}^{\boldsymbol{G}}(\mu_{\mathfrak{n}},\mathfrak{c})^{\mathrm{ord}}_{\mathcal{O}_E},\omega(k,l)\right)$.

Recall also, from section 4.5, that the line bundles $\omega(k,l)$ of interest when dealing with Hilbert modular forms of weight (v,w) are given by taking $k_\tau=v_\tau$, $l_\tau=\frac{w-v_\tau}{2}$. In these terms, the normalisation performed above thus yields integrality of the Hecke operator

$$\mathrm{T}_{\mathfrak{p}}=\left(\mathbf{N}^F_{\mathbb{Q}}\mathfrak{p}\right)^{-1}\cdot\prod_\tau\tau(x)^{\frac{v_\tau-w}{2}}\cdot\mathrm{T}^{\mathrm{naive}}_{\mathfrak{p}}$$

acting on Hilbert modular forms of weight (v,w).

5.3 Integrality of Hecke operators at primes dividing the level

We now want to study the integrality property of Hecke operators acting on spaces of Hilbert modular forms at primes dividing the level. To do this, we use the q-expansion principle (Theorem 4.10): to study the integrality properties of Hecke operators, it is enough to study the effect that such a Hecke operator has on integrality of q-expansions.

We return to the diagrams of Hecke operators defining Hecke correspondences :

$$\begin{array}{ccc}
\mathrm{H}^0\left(M(U_0(\mathfrak{n}),\mathfrak{p}\mathfrak{c})_E,\omega(\kappa)\right) & \xrightarrow{(p_2)^*} & \mathrm{H}^0\left(Y(U_0(\mathfrak{n}),\mathfrak{c},\mathfrak{p}\mathfrak{c})_E,(p_2)^*\omega(\kappa)\right) \\
& & \Big\downarrow{\scriptstyle\alpha^*} \\
& & \mathrm{H}^0\left(Y(U_0(\mathfrak{n}),\mathfrak{c},\mathfrak{p}\mathfrak{c})_E,(p_1)^*\omega(\kappa)\right) \xrightarrow[\mathrm{tr}(p_1)]{} \mathrm{H}^0\left(M(U_0(\mathfrak{n}),\mathfrak{c})E,\omega(\kappa)\right).
\end{array}$$

We start with the Tate abelian variety around a cusp labelled by $(\mathfrak{a},\mathfrak{b})$, $\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q)$, as defined in section 4.7.2. To compute the effect of the Hecke operator $\mathrm{U}_{\mathfrak{p}}$ on q-expansions, we then need to know the relevant isogenies from $\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q)$.[81] Recall from section 4.7.2 that the $\mathfrak{p}$-torsion of $\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q)$ fits naturally in the following split exact sequence

$$0\longrightarrow\mu_{\mathfrak{p}}\longrightarrow\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q)[\mathfrak{p}]\longrightarrow\mathfrak{p}^{-1}\mathfrak{b}/\mathfrak{b}\longrightarrow 0,$$

where we assumed without loss of generality that $\mathfrak{p}$ was coprime to $\mathfrak{a}$.

The rank 1 $\mathcal{O}_F/\mathfrak{p}$-module subschemes of the $\mathfrak{p}$-torsion are then of two forms:

- $\mu_{\mathfrak{p}}\hookrightarrow\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q)[\mathfrak{p}]$,
- $\{(Q\zeta)^i\mid i\in\mathcal{O}_F/\mathfrak{p}\}$ for $\zeta\in\mu_{\mathfrak{p}}$ and $Q=q^x$ for some $x\in F$ with $\mathfrak{p}^{-1}\mathfrak{b}=\mathfrak{b}+x\mathcal{O}_F$ (i.e. Q is a "$\mathfrak{p}$-th root of q").

The corresponding quotients are then of the following form:

- $\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q)/\mu_{\mathfrak{p}}\cong\mathrm{Tate}_{\mathfrak{p}\mathfrak{a},\mathfrak{b}}(q)$ induced by the natural morphism $\mathbb{G}_{\mathrm{m}}\otimes_{\mathbb{Z}}\mathfrak{a}^\vee\to\mathbb{G}_{\mathrm{m}}\otimes_{\mathbb{Z}}\mathfrak{p}^{-1}\mathfrak{a}^\vee$, and
- $\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q)/\{(Q\zeta)^i\mid i\in\mathcal{O}_F/\mathfrak{p}\}\cong\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(Q\zeta)$.

Write the corresponding isogeny $\alpha_\zeta\colon\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q)\to\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(Q\zeta)$. Note that $\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(Q\zeta)$ is $\mathfrak{p}\mathfrak{a}\mathfrak{b}^{-1}$-polarised, and not $\mathfrak{a}\mathfrak{b}^{-1}$-polarised.

[81] Note that we are interested in isogenies *from* $\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q)$ and not *to* it. This is because the pullback by the universal isogeny, a key ingredient in the definition of the Hecke operators, forces us to consider the different abelian varieties under a given abelian variety, and not the other way around.

The next port of call is to study the effect of α_ζ on the level of the line bundles $\omega_{\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q)}$. To do this, we use the canonical differential top-forms on $\mathbb{G}_\mathrm{m} \otimes_\mathbb{Z} \mathfrak{a}^\vee$; the isogeny α_ζ then fits into the following diagram:

$$\begin{tikzcd} \mathbb{G}_\mathrm{m} \otimes_\mathbb{Z} \mathfrak{a}^\vee \arrow[r, "\mathrm{id}"] \arrow[d] & \mathbb{G}_\mathrm{m} \otimes_\mathbb{Z} \mathfrak{a}^\vee \arrow[d] \\ \mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q) \arrow[r, "\alpha_\zeta"] & \mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(Q\zeta) \end{tikzcd}$$

As a consequence, $(\alpha_\zeta)^*$ is the identity on the level of canonical differentials.

As in [Hida. p-adic forms, §4.2.9], we consider the Katz description of modular forms. Letting u be the canonical differential of $\omega_{\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q)}$, and leaving the $\mu_\mathfrak{n}$-level structures implicit, the effect of the (non-normalised) Hecke operator $\mathrm{U}_\mathfrak{p}^{\mathrm{naive}}$ is thus:

$$\begin{aligned}(\mathrm{U}_\mathfrak{p}^{\mathrm{naive}} f)(\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q), u) &= \sum_{\zeta\in\mu_\mathfrak{p}} f(\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(\zeta Q), (\alpha_\zeta)^* u) \\ &= \sum_{\zeta\in\mu_\mathfrak{p}} f(\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(\zeta Q), u) \\ &= \sum_{\zeta\in\mu_\mathfrak{p}} \sum_{\xi\in(\mathfrak{a}\mathfrak{b})_+\cup\{0\}} \mathrm{a}_{(\mathfrak{a},\mathfrak{b})}(f,\xi)(\zeta Q)^\xi.\end{aligned}$$

Now, by orthogonality of the characters $\zeta \mapsto \zeta^\xi$ of $\mu_\mathfrak{p}$, the contributions for $\xi \in (\mathfrak{a}\mathfrak{b})_+ \smallsetminus (\mathfrak{p}\mathfrak{a}\mathfrak{b})_+$ cancel. Hence we can perform the sum over $\xi \in (\mathfrak{p}\mathfrak{a}\mathfrak{b})_+ \cup \{0\}$ instead. Finishing the calculation, we then have:

$$\begin{aligned}(\mathrm{U}_\mathfrak{p}^{\mathrm{naive}} f)(\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q), u) &= \sum_{\zeta\in\mu_\mathfrak{p}} \sum_{\xi\in(\mathfrak{p}\mathfrak{a}\mathfrak{b})_+\cup\{0\}} \mathrm{a}_{(\mathfrak{a},\mathfrak{b})}(f,\xi)Q^\xi \\ &= \mathbf{N}_\mathbb{Q}^F(\mathfrak{p}) \sum_{\xi\in(\mathfrak{p}\mathfrak{a}\mathfrak{b})_+\cup\{0\}} \mathrm{a}_{(\mathfrak{a},\mathfrak{b})}(f,\xi)Q^\xi.\end{aligned}$$

which, as required, is a multiple of $\mathbf{N}_\mathbb{Q}^F(\mathfrak{p})$. We thus conclude:

Proposition 5.3 — The Hecke operator

$$\mathrm{U}_\mathfrak{p}^{\mathrm{naive}}\colon \mathrm{H}^0\left(M(U_0(\mathfrak{n}), \mathfrak{p}\mathfrak{c})_E, \omega(\kappa)\right) \longrightarrow \mathrm{H}^0\left(M(U_0(\mathfrak{n}), \mathfrak{c})E, \omega(\kappa)\right)$$

extends, by the q-expansion principle, to an operator on spaces of integral Hilbert modular forms (for $\boldsymbol{G}^*$)

$$\mathrm{U}_\mathfrak{p}^{\mathrm{naive}}\colon \mathrm{M}_\kappa^{\boldsymbol{G}^*}(\mu_\mathfrak{n}, \mathfrak{p}\mathfrak{c}; \mathcal{O}_E) \longrightarrow \mathrm{M}_\kappa^{\boldsymbol{G}^*}(\mu_\mathfrak{n}, \mathfrak{c}; \mathcal{O}_E),$$

with image contained in $\mathbf{N}_\mathbb{Q}^F\mathfrak{p} \cdot \mathrm{M}_\kappa^{\boldsymbol{G}^*}(\mu_\mathfrak{n}, \mathfrak{c}; \mathcal{O}_E)$. ◇□

Using the method of proof of Proposition 5.2, we then deduce the corresponding result for $\boldsymbol{G}$:

Proposition 5.4 — The Hecke operator

$$\mathrm{U}_\mathfrak{p}^{\mathrm{naive}}\colon \mathrm{H}^0\left(M^{\boldsymbol{G}}(U_0(\mathfrak{n}), \mathfrak{p}\mathfrak{c})_E, \omega(k,l)\right) \longrightarrow \mathrm{H}^0\left(M^{\boldsymbol{G}}(U_0(\mathfrak{n}), \mathfrak{c})E, \omega(k,l))\right)$$

extends, by the q-expansion principle, to an operator on spaces of integral Hilbert modular forms (for $\boldsymbol{G}$)

$$\mathrm{U}_\mathfrak{p}^{\mathrm{naive}}\colon \mathrm{M}_{(v,w)}^{\boldsymbol{G}}(\mu_\mathfrak{n}, \mathfrak{p}\mathfrak{c}; \mathcal{O}_E) \longrightarrow \mathrm{M}_{(v,w)}^{\boldsymbol{G}}(\mu_\mathfrak{n}, \mathfrak{c}; \mathcal{O}_E),$$

with image contained in $\mathbf{N}_\mathbb{Q}^F\mathfrak{p} \cdot \prod_{\tau:F\hookrightarrow E} \tau(\mathfrak{p})^{\frac{v_\tau - w}{2}} \cdot \mathrm{M}_\kappa^{\boldsymbol{G}^*}(\mu_\mathfrak{n}, \mathfrak{c}; \mathcal{O}_E)$. ◇□

Finally, we can also eliminate polarisations as explained in section 4.5, to obtain a Hecke operator

$$\mathrm{U}_\mathfrak{p}^{\mathrm{naive}}\colon \mathrm{M}_{(v,w)}^{\boldsymbol{G}}(\mu_\mathfrak{n}; \mathcal{O}_E) \longrightarrow \mathrm{M}_{(v,w)}^{\boldsymbol{G}}(\mu_\mathfrak{n}; \mathcal{O}_E),$$

with the same property.

6 Semistability versus weights, part II

The integrality result for Hecke operators will now allow us to rule out the local components of the automorphic representation attached to a low weight Hilbert modular form from being special.[82]

This mirrors Theorem I.14.4, which imposed a condition on the Hodge–Tate weights for semistable non-crystalline filtered (φ, N)-modules (such as those arising from Galois representations attached to Hilbert modular forms).

6.1 Splitting behaviour versus embeddings

Recall from Proposition 5.4 that the image of

$$\mathrm{U}_{\mathfrak{p}}^{\mathrm{naive}}\colon \mathrm{M}_{(\nu,w)}^{\boldsymbol{G}}(\mu_{\mathfrak{n}}, \mathfrak{p}\mathfrak{c}; \mathcal{O}_E) \longrightarrow \mathrm{M}_{(\nu,w)}^{\boldsymbol{G}}(\mu_{\mathfrak{n}}, \mathfrak{c}; \mathcal{O}_E)$$

is contained in $\mathbf{N}_{\mathbb{Q}}^{F}\mathfrak{p} \cdot \prod_{\tau\colon F\hookrightarrow E} \tau(\mathfrak{p})^{\frac{\nu_\tau - w}{2}} \cdot \mathrm{M}_{(\nu,w)}^{\boldsymbol{G}}(\mu_{\mathfrak{n}}, \mathfrak{c}; \mathcal{O}_E)$.

Considering the situation locally at $\mathfrak{p}$, we can choose a uniformiser ϖ of the completion $F_{\mathfrak{p}}$. Then $\prod_\tau \tau^{\nu_\tau - w}$ can be seen as a character of $F_{\mathfrak{p}}$ by continuity. To understand the integrality properties of $\mathrm{U}_{\mathfrak{p}}^{\mathrm{naive}}$ we are thus lead to study the integrality of $\chi(\varpi)$ for arbitrary characters χ. A basic lemma will allow us to get a handle on the situation.

The splitting behaviour of p in $\mathcal{O}_F$ determines various collections of subsets of the set of embeddings $\Sigma = \{\tau\colon F \hookrightarrow \mathbb{C}\}$, corresponding to the different primes above p. Choosing a Galois closure E/F of F, and given $\mathfrak{p} \mid p$, we define a collection of subsets $P_{\mathfrak{p}}$ of Σ, which are meant to keep track of where $\mathfrak{p}$ is mapped to, relative to the primes $\mathfrak{q}$ of E above p.

- Write $G = \mathrm{Gal}(E/\mathbb{Q})$, $H = \mathrm{Gal}(E/F) \leqslant G$. A choice of embedding $\nu\colon E \to \mathbb{C}$ gives a bijection $G/H \to \Sigma$, $[g] \mapsto (\nu \circ g)|_F$.
- Pick a prime $\mathfrak{q} \subset \mathcal{O}_E$ above $\mathfrak{p}$.
- Then $P_{\mathfrak{p}}$ is the collection of subsets of Σ determined by the collection of subsets of G/H given by $\{(g\mathrm{D}_{\mathfrak{q}}H)/H \mid g \in G\}$, through the choice of embedding $\nu\colon E \to \mathbb{C}$.

Here $\mathrm{D}_{\mathfrak{q}}$ is the decomposition subgroup at $\mathfrak{q}$. Note that $P_{\mathfrak{p}}$ is independent of E, of ν, and of the choice of $\mathfrak{q}$ above $\mathfrak{p}$, as given $\mathfrak{q}_1, \mathfrak{q}_2$ above $\mathfrak{p}$, there always exists $h \in H$ such that $h(\mathfrak{q}_1) = \mathfrak{q}_2$.

In the case that F is Galois, this simplifies to $P_{\mathfrak{p}}$ being the partition of Σ given by left $\mathrm{D}_{\mathfrak{q}}$-cosets.

We can now state the aforementioned lemma:

Lemma 6.1 — Consider a character $\chi\colon F \to E$ given by $\chi = \prod_{\tau\in\Sigma} \tau^{n_\tau}$, for some $n_\tau \in \mathbb{Z}$.
Let $\mathfrak{p} \mid p$ be a prime ideal of $\mathcal{O}_F$, and $x \in \mathfrak{p}$ an element such that no other prime above p divides x.
Then $\chi(x)$ is a p-adic algebraic integer if and only if for every subset $C \subseteq \Sigma$ from $P_{\mathfrak{p}}$, we have that $\sum_{\tau\in C} n_\tau \geqslant 0$. ◇

Proof: The claim is local at p, so we start by completing. Completing E at p yields the product $E_p = \prod_{\mathfrak{q}|p} E_{\mathfrak{q}}$.
We can then check p-integrality by checking integrality separately in each $E_{\mathfrak{q}}$.
Now, using the assumption that no other primes above p divide x, we can assume that $x = \varpi^r$ for some uniformiser ϖ of $F_{\mathfrak{p}}$. In fact, we can also assume that $r = 1$, as integrality is unchanged under taking positive powers.
For each embedding $\tau\colon F \hookrightarrow E$, we then consider the completions $E_{\mathfrak{p},\tau} := \prod_{\mathfrak{q}|(\tau(\mathfrak{p}))} E_{\mathfrak{q}}$. Varying τ, we collect embeddings together when $E_{\mathfrak{p},\tau}$ share one factor $E_{\mathfrak{q}}$. Using the action of G to vary τ, we see that these subsets are precisely the subsets of the form $\{(g\mathrm{D}_{\mathfrak{q}}H)/H\}$ that define $P_{\mathfrak{p}}$.
Given a subset $C \in P_{\mathfrak{p}}$, the valuations of $\tau(\varpi)$ in the common factors $E_{\mathfrak{q}}$ of the $E_{\mathfrak{p},\tau}$ are independent of τ, so that integrality of $\chi(\tau)$ for these factors is equivalent to $\sum_C n_\tau \geqslant 0$. □

[82] Recall that a smooth admissible representation of a p-adic group is called special if it is a discrete series which is not supercuspidal; in the case of $\mathrm{GL}_2(K)$ this just means that it is a twist of the Steinberg representation.

6.2 The weight condition

We can now state the automorphic analogue of Theorem I.14.4, for Hilbert modular forms.

Theorem 6.2 — Let π be the automorphic representation attached to a Hilbert modular form of weight $((\nu_\tau)_\tau, w)$ over F. Let $\mathfrak{p}$ be a prime above p in $\mathcal{O}_F$, and $P_\mathfrak{p}$ the the collection of subsets of Σ associated with $\mathfrak{p}$.
If $\pi_\mathfrak{p}$ is special, then the weights ν_τ average at least 2 over every subset in $P_\mathfrak{p}$. $\diamond$

Proof: We compare the known action of Hecke operators on special representations given by Proposition 3.3 with the integrality result for the operator $\mathrm{U}_\mathfrak{p}^{\text{naive}}$ of Proposition 5.4.
We begin by choosing $\varpi \in F$ that uniformises $\mathfrak{p}$ and that is divisible by no other prime ideal of $\mathcal{O}_F$ above p.
Let then $F_\mathfrak{p}$ be the completion of F at $\mathfrak{p}$, uniformised by ϖ. The classical double coset operator $\mathrm{T}_\varpi^{\text{naive}}$ of $\mathrm{GL}_2(F_\mathfrak{p})$ acts by $q := q_\varpi = \mathbf{N}_\mathbb{Q}^F \mathfrak{p}$ on the Steinberg representation St, as shown by Proposition 3.3.
Now, assume that $\pi_\mathfrak{p}$ is special. We know that the local component $\pi_\mathfrak{p}$ is isomorphic to

$$\mathrm{St} \otimes \left(\left(\chi_f \otimes \|-\|^{\frac{d(2-w)}{2}} \right) \circ \det \right),$$

with χ_f a finite order character. Thus, up to a unit, $\mathrm{T}_\varpi^{\text{naive}}$ acts on $\pi_\mathfrak{p}$ by $q \cdot \prod_\tau \tau(\varpi)^{\frac{w-2}{2}}$.
On the other hand, Proposition 5.4 implies that the operator $q^{-1} \cdot \prod_\tau \tau(\varpi)^{\frac{\nu_\tau - w}{2}} \mathrm{U}_\mathfrak{p}^{\text{naive}}$ is p-adically integral. Note that we are using the assumption that no other prime of $\mathcal{O}_F$ above p divides ϖ to perform the division by $\prod_\tau \tau(\varpi)^{\frac{w-\nu_\tau}{2}}$. Consequently, the quantity

$$\prod_\tau \tau(\varpi)^{\frac{\nu_\tau - w}{2}} \prod_\tau \tau(\varpi)^{\frac{w-2}{2}} = \prod_\tau \tau(\varpi)^{\frac{\nu_\tau - 2}{2}}$$

must be p-adically integral. Applying Lemma 6.1, we obtain the desired result. $\square$

6.2.1 Examples

We can be more concrete about the implications of this theorem in various situations.
For instance, if the prime p is inert, the condition on the weight of $\pi_{(p)}$ is the inequality $\sum_\tau \nu_\tau \geqslant 2d$, which mirrors the inequality from section I.14.
On the opposite side of the spectrum, if p splits completely, then we get one condition for each embedding: $\nu_\tau \geqslant 2$ for every $\tau \in \Sigma$.
More interesting behaviour can occur when the splitting behaviour is more complicated. Consider for instance the non-Galois totally real cubic field $F = \mathbb{Q}[x]/(x^3 - 4x - 1)$, with prime discriminant 229. Write $\alpha = [x] \in F$. Let E be the Galois closure of F; it is the compositum $E = F \cdot \mathbb{Q}(\sqrt{229})$. We have that $\mathrm{Gal}(E/\mathbb{Q}) = \langle r, s | r^3 = 1, s^2 = 1, (sr)^2 = 1 \rangle \cong \mathrm{S}_3$, where:

$$s\colon \sqrt{229} \mapsto -\sqrt{229}, \qquad r\colon \sqrt{229} \mapsto \sqrt{229},$$

$$s\colon \alpha \mapsto \alpha, \qquad r\colon \alpha \mapsto \frac{1}{2}\left(\frac{\sqrt{229}}{229}\left(24\alpha^2 - 9\alpha - 64\right) - \alpha \right).$$

The splitting behaviour of a rational prime $p \in \mathbb{Z}$ in the tower $E/F/\mathbb{Q}$ then falls into one of four categories,[83] the most interesting of which is the following:

$$\begin{tikzcd}
E \arrow[d, dash] & & sr(\mathfrak{p})\mathcal{O}_E \arrow[dr, dash] & & r(\mathfrak{p})\mathcal{O}_E \arrow[dl, dash] & & \mathfrak{p}\mathcal{O}_E \arrow[dl, dash] \\
F \arrow[d, dash] & & & \mathfrak{q} \arrow[dr, dash] & & \mathfrak{p} \arrow[dl, dash] & \\
\mathbb{Q} & & & & p & &
\end{tikzcd}$$

[83] Either $p = 229$, or p is unramified in which case there are 3 possibilities, corresponding to the 3 conjugacy classes of cyclic subgroups of S_3.

The primes p exhibiting this splitting behaviour are precisely those satisfying $\left(\frac{229}{p}\right) = -1$,[84] such as $2, 7, 13, 23, 29, 31, \ldots$.
In the notation of section 6.1, we then have $G := \mathrm{Gal}(E/\mathbb{Q}) = \langle r, s \,|\, r^3 = 1, s^2 = 1, (sr)^2 = 1\rangle$, $H = \langle s\rangle \subset G$, $\mathrm{D}_{\mathfrak{p}\mathcal{O}_E} = \langle s\rangle \subset G$, $\mathrm{D}_{r(\mathfrak{p})\mathcal{O}_E} = \langle rsr^{-1}\rangle = \langle sr\rangle \subset G$. As $\mathrm{D}_{\mathfrak{p}\mathcal{O}_E} H = H$, $P_\mathfrak{p}$ is the set of singletons in G/H. On the other hand, $\mathrm{D}_{r(\mathfrak{p})\mathcal{O}_E} H = \{1, s, sr, r^2\}$, so that $P_\mathfrak{q}$ consists of all 2-element subsets of G/H.
As a consequence, Theorem 6.2 yields two distinct conditions on the weight $((\nu_\tau)_\tau, w)$ of a Hilbert modular f form over F:

- if π_f is special at $\mathfrak{p}$, we must have that $\nu_\tau \geqslant 2$ for all $\tau \in \Sigma$,
- if π_f is special at $\mathfrak{q}$, we must have that $\frac{1}{2}(\nu_{\tau_1} + \nu_{\tau_2}) \geqslant 2$ for all pairs $\tau_1, \tau_2 \in \Sigma$, $\tau_1 \neq \tau_2$.

For instance, if $(\nu_\tau)_\tau = (1, 1, 5)$, then π_f cannot be special at $\mathfrak{q}$ even though $\frac{1+1+5}{3} > 2$, but one cannot exclude a form with $(\nu_\tau)_\tau = (1, 3, 3)$ from having a special local component at $\mathfrak{q}$.

To recapitulate then, Theorem 6.2 shows that the local components of the automorphic representation π_f attached to a Hilbert modular form f cannot be special if the weights ν_τ of f do not average at least 2 over certain subsets $C \subseteq \{\tau\colon F \hookrightarrow \mathbb{R}\}$. In particular, this is only interesting for Hilbert modular forms of partial weight 1 (or parallel weight 1). One source of such Hilbert modular forms is given by applying automorphic induction to certain Hecke characters of a CM extension of F [Moy–Specter, §2.5]. This is not very interesting, as such forms are everywhere potentially unramified, so Theorem 6.2 tells us nothing new. However, these are not the only Hilbert modular forms of partial weight 1: an example of a Hilbert modular form f of weight $(1, 5)$ over the quadratic field $F = \mathbb{Q}(\sqrt{5})$ is given in [Moy–Specter, Theorem 3.1]. Moreover, it is shown that the local component $\pi_{f,(7)}$ is special. Noting that 7 is inert in $\mathbb{Q}(\sqrt{5})$, this is indeed compatible with the above theorem, because

$$\frac{1+5}{2} = 3 \geqslant 2.$$

[84] Indeed, if $p \neq 229$, $\left(\frac{229}{p}\right) = (-1)^{6-g} = (-1)^g$, where g denotes the number of primes above p in E. This interesting case corresponds to the unique situation in which g is odd.

Bibliography for part II

[Andreatta–Goren] Fabrizio Andreatta and Eyal Zvi Goren. “Hilbert modular forms: mod p and p-adic aspects”. In: *Mem. Amer. Math. Soc.* 173.819 (2005), pp. vi+100.

[Andreatta–Iovita–Pilloni] Fabrizio Andreatta, Adrian Iovita, and Vincent Pilloni. “On overconvergent Hilbert modular forms”. In: *Astérisque* 382 (2016), pp. 163–193.

[A–M–R–T] Avner Ash, David Mumford, Michael Rapoport, and Yung-Sheng Tai. *Smooth compactifications of locally symmetric varieties*. Second edition. Cambridge Mathematical Library. With the collaboration of Peter Scholze. Cambridge University Press, Cambridge, 2010, pp. x+230.

[Blasius–Rogawski] Don Blasius and Jonathan Rogawski. “Galois representations for Hilbert modular forms”. In: *Bull. Amer. Math. Soc. (N.S.)* 21.1 (1989), pp. 65–69.

[Breuil–Mézard] Christophe Breuil and Ariane Mézard. “Multiplicités modulaires et représentations de $\mathrm{GL}_2(\mathbb{Z}_p)$ et de $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ en $\ell = p$”. In: *Duke Math. J.* 115.2 (2002). With an appendix by Guy Henniart, pp. 205–310.

[Carayol. $\rho_{f,\ell}$] Henri Carayol. “Sur les représentations ℓ-adiques associées aux formes modulaires de Hilbert”. In: *Ann. Sci. École Norm. Sup. (4)* 19.3 (1986), pp. 409–468.

[Casselman. Unramified P.S. I] Bill Casselman. “The unramified principal series of p-adic groups. I. The spherical function”. In: *Compositio Math.* 40.3 (1980), pp. 387–406.

[Chai. Compactifications] Ching-Li Chai. “Arithmetic minimal compactification of the Hilbert-Blumenthal moduli spaces”. In: *Ann. of Math. (2)* 131.3 (1990), pp. 541–554.

[Conrad. Moduli] Brian Conrad. “Arithmetic moduli of generalized elliptic curves”. In: *J. Inst. Math. Jussieu* 6.2 (2007), pp. 209–278.

[Deligne. $\rho_{f,\ell}$] Pierre Deligne. “Formes modulaires et représentations ℓ-adiques”. In: *Séminaire Bourbaki. Vol. 1968/69: Exposés 347–363*. Vol. 175. Lecture Notes in Math. Springer, Berlin, 1971, Exp. No. 355, 139–172.

[Deligne. Letter to P-S] Pierre Deligne. “Letter to Piatetski-Shapiro”. Mar. 1973.

[Deligne. Hodge III] Pierre Deligne. “Théorie de Hodge. III”. In: *Inst. Hautes Études Sci. Publ. Math.* 44 (1974), pp. 5–77.

[Deligne. Shimura] Pierre Deligne. “Travaux de Shimura”. In: *Séminaire Bourbaki, 23ème année (1970/71), Exp. No. 389*. Springer, Berlin, 1971, 123–165. Lecture Notes in Math., Vol. 244.

[Deligne–Pappas] Pierre Deligne and Georgios Pappas. “Singularités des espaces de modules de Hilbert, en les caractéristiques divisant le discriminant”. In: *Compositio Math.* 90.1 (1994), pp. 59–79.

[Deligne–Rapoport] Pierre Deligne and Michael Rapoport. “Les schémas de modules de courbes elliptiques”. In: *Modular functions of one variable, II (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972)*. Springer, Berlin, 1973, 143–316. Lecture Notes in Math., Vol. 349.

[Deligne–Serre] Pierre Deligne and Jean-Pierre Serre. “Formes modulaires de poids 1”. In: *Ann. Sci. École Norm. Sup. (4)* 7 (1974), 507–530 (1975).

[Dimitrov–Tilouine] Mladen Dimitrov and Jacques Tilouine. “Variétés et formes modulaires de Hilbert arithmétiques pour $\Gamma_1(c, n)$”. In: *Geometric aspects of Dwork theory. Vol. I, II*. Walter de Gruyter GmbH & Co. KG, Berlin, 2004, pp. 555–614.

[Emerton–Reduzzi–Xiao] Matthew Emerton, Davide Reduzzi, and Liang Xiao. *Unramifiedness of Galois representations arising from Hilbert modular surfaces*. Preprint. May 2016. URL: http://arxiv.org/abs/1410.6203.

[Faltings–Chai] Gerd Faltings and Ching-Li Chai. *Degeneration of abelian varieties*. Vol. 22. Ergebnisse der Mathematik und ihrer Grenzgebiete (3). With an appendix by David Mumford. Springer-Verlag, Berlin, 1990, pp. xii+316.

[Hida. p-adic Hecke algebras] Haruzo Hida. "On p-adic Hecke algebras for GL_2 over totally real fields". In: *Ann. of Math. (2)* 128.2 (1988), pp. 295–384.

[Hida. p-adic forms] Haruzo Hida. *p-adic automorphic forms on Shimura varieties*. Springer Monographs in Mathematics. Springer-Verlag, New York, 2004, pp. xii+390.

[Iwahori–Matsumoto] Nagayoshi Iwahori and Hideya Matsumoto. "On some Bruhat decomposition and the structure of the Hecke rings of p-adic Chevalley groups". In: *Inst. Hautes Études Sci. Publ. Math.* 25 (1965), pp. 5–48.

[Jarvis. $\rho_{f,\ell}$] Frazer Jarvis. "On Galois representations associated to Hilbert modular forms". In: *J. Reine Angew. Math.* 491 (1997), pp. 199–216.

[Kazhdan–Lusztig] David Kazhdan and George Lusztig. "Proof of the Deligne-Langlands conjecture for Hecke algebras". In: *Invent. Math.* 87.1 (1987), pp. 153–215.

[K–K–M–S-D] George Kempf, Finn Faye Knudsen, David Mumford, and Bernard Saint-Donat. *Toroidal embeddings. I*. Lecture Notes in Mathematics, Vol. 339. Springer-Verlag, Berlin-New York, 1973, pp. viii+209.

[Kisin–Lai] Mark Kisin and King Fai Lai. "Overconvergent Hilbert modular forms". In: *Amer. J. Math.* 127.4 (2005), pp. 735–783.

[Lansky. Double cosets] Joshua M. Lansky. "Decomposition of double cosets in p-adic groups". In: *Pacific J. Math.* 197.1 (2001), pp. 97–117.

[Loeffler–Weinstein] David Loeffler and Jared Weinstein. "On the computation of local components of a newform". In: *Math. Comp.* 81.278 (2012), pp. 1179–1200.

[Moy–Specter] Richard Moy and Joel Specter. "There exist non-CM Hilbert modular forms of partial weight 1". In: *Int. Math. Res. Not. IMRN* 24 (2015), pp. 13047–13061.

[Rapoport. Compactifications] Michael Rapoport. "Compactifications de l'espace de modules de Hilbert-Blumenthal". In: *Compositio Math.* 36.3 (1978), pp. 255–335.

[T. Saito. HMF] Takeshi Saito. "Hilbert modular forms and p-adic Hodge theory". In: *Compos. Math.* 145.5 (2009), pp. 1081–1113.

[Scholl. Motives] Tony Scholl. "Motives for modular forms". In: *Invent. Math.* 100.2 (1990), pp. 419–430.

[SGA 7_{I}] *Groupes de monodromie en géométrie algébrique. I*. Lecture Notes in Mathematics, Vol. 288. Séminaire de Géométrie Algébrique du Bois-Marie 1967–1969 (SGA 7). Dirigé par Alexander Grothendieck. Avec la collaboration de Michèle Raynaud et Dock Sang Rim. Springer-Verlag, Berlin-New York, 1972, pp. viii+523.

[Skinner. Note] Christopher Skinner. "A note on the p-adic Galois representations attached to Hilbert modular forms". In: *Doc. Math.* 14 (2009), pp. 241–258.

[Stroh. Compactifications] Benoît Stroh. "Compactification de variétés de Siegel aux places de mauvaise réduction". In: *Bull. Soc. Math. France* 138.2 (2010), pp. 259–315.

[Taylor. $\rho_{f,\ell}$] Richard Taylor. "On Galois representations associated to Hilbert modular forms". In: *Invent. Math.* 98.2 (1989), pp. 265–280.

[Zelevinsky. Segments] Andrei Zelevinsky. "Induced representations of reductive p-adic groups. II. On irreducible representations of $\mathrm{GL}(n)$". In: *Ann. Sci. École Norm. Sup. (4)* 13.2 (1980), pp. 165–210.

Conclusion

As explained in the introduction, the construction of Galois representations associated with Hilbert modular forms is only partly geometric. Indeed, some of these Galois representations occur (via the Jacquet–Langlands correspondence) in the cohomology of certain quaternionic Shimura curves, whereas in other situations (the *non-cohomological case*), these Galois representations cannot be found, in this way, within the cohomology of any Shimura variety. The missing Galois representations can instead be constructed by congruences, but this procedure does not keep track of important data, such as the precise action of the monodromy operator N. As an example, for the non-CM Hilbert modular form f over $F = \mathbb{Q}(\sqrt{5})$ of weight $(1, 5)$ mentioned at the end of section II.6.2.1, which is special at 7, it is difficult to ascertain whether the monodromy operator N is nontrivial in the Weil–Deligne representations $\mathrm{WD}(\rho_{f,\ell,(7)})$, or whether $\rho_{f,7,(7)}$ is semistable (non-crystalline) with expected Hodge–Tate weights. This is the problem of *local-global compatibility*.

The parallel results of Theorem I.14.4 and Theorem II.6.2 can be seen as evidence of local-global compatibility for Hilbert modular forms of partial weight one, as they give identical conditions, respectively on the geometric and automorphic sides, which enforce vanishing of monodromy.